

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 3, 2004, 09:12:48 ; Search time 34 Seconds  
(without alignments)  
1607.317 Million cell updates/sec

Title: US-09-990-415A-2  
Perfect score: 1345  
Sequence: 1 MESKMGELPLDINIOEPRWD.....PAMAIPLIMDTLEKKDFLK 261

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 801455 seqs, 209382283 residues

Total number of hits satisfying chosen parameters: 801455

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1345	100.0	261	US-09-990-415A-2	Sequence 2, Appl1
2	1345	100.0	266	US-09-867-550-1344	Sequence 1344, Ap
3	1089	81.0	322	US-09-990-415A-8	Sequence 8, Appl1
4	1089	81.0	322	US-10-014-338-2	Sequence 2, Appl1
5	876	65.1	261	US-10-014-338-4	Sequence 4, Appl1
6	753	56.0	322	US-09-990-415A-6	Sequence 6, Appl1
7	436	32.4	186	US-10-094-749-2165	Sequence 2165, Ap
8	368.5	27.4	77	US-09-990-415A-4	Sequence 4, Appl1
9	312	23.2	127	US-10-264-237-2468	Sequence 2468, Ap
10	143.5	10.7	337	US-10-415-378-1	Sequence 1, Appl1
11	143	10.6	54	US-10-106-698-6940	Sequence 6940, Ap
12	96	7.1	836	US-09-972-708-9	Sequence 9, Appl1
13	91	6.8	496	US-10-104-047-3006	Sequence 3006, Ap
14	89.5	6.7	344	US-10-306-133-3	Sequence 3, Appl1
15	88.5	6.6	344	US-09-978-295A-523	Sequence 523, App

16	88.5	6.6	344	10	US-09-966-546-4	Sequence 4, Appl1
17	88.5	6.6	344	10	US-09-966-546-6	Sequence 6, Appl1
18	88.5	6.6	344	10	US-09-978-697-523	Sequence 523, App
19	88.5	6.6	344	10	US-09-966-545-4	Sequence 4, Appl1
20	88.5	6.6	344	10	US-09-966-545-6	Sequence 6, Appl1
21	88.5	6.6	344	10	US-09-978-192A-523	Sequence 523, App
22	88.5	6.6	344	10	US-09-999-832A-523	Sequence 523, App
23	88.5	6.6	344	11	US-09-965-212-4	Sequence 4, Appl1
24	88.5	6.6	344	11	US-09-965-212-6	Sequence 6, Appl1
25	88.5	6.6	344	11	US-09-978-189-523	Sequence 523, App
26	88.5	6.6	344	11	US-09-978-608A-523	Sequence 523, App
27	88.5	6.6	344	11	US-09-978-585A-523	Sequence 523, App
28	88.5	6.6	344	11	US-09-978-191A-523	Sequence 523, App
29	88.5	6.6	344	11	US-09-978-403A-523	Sequence 523, App
30	88.5	6.6	344	11	US-09-978-564A-523	Sequence 523, App
31	88.5	6.6	344	11	US-09-999-833A-523	Sequence 523, App
32	88.5	6.6	344	11	US-09-981-915A-523	Sequence 523, App
33	88.5	6.6	344	11	US-09-978-824-523	Sequence 523, App
34	88.5	6.6	344	11	US-09-918-585A-523	Sequence 523, App
35	88.5	6.6	344	11	US-09-978-423A-523	Sequence 523, App
36	88.5	6.6	344	11	US-09-978-193A-523	Sequence 523, App
37	88.5	6.6	344	11	US-09-999-830A-523	Sequence 523, App
38	88.5	6.6	344	11	US-09-978-757A-523	Sequence 523, App
39	88.5	6.6	344	11	US-09-978-187B-523	Sequence 523, App
40	88.5	6.6	344	11	US-09-978-643A-523	Sequence 523, App
41	88.5	6.6	344	12	US-09-978-375A-523	Sequence 523, App
42	88.5	6.6	344	12	US-09-978-188A-523	Sequence 523, App
43	88.5	6.6	344	12	US-09-978-298A-523	Sequence 523, App
44	88.5	6.6	344	12	US-10-137-870-376	Sequence 376, App
45	88.5	6.6	344	12	US-10-140-018-376	Sequence 376, App

ALIGNMENTS

RESULT 1  
US-09-990-415A-2  
; Sequence 2, Application US/0990415A  
; Patent No. US20020165182A1  
; GENERAL INFORMATION:  
; APPLICANT: Pharmacia AB  
; TITLE OF INVENTION: Protein Cluster I  
; FILE REFERENCE: 00349  
; CURRENT APPLICATION NUMBER: US/09/990,415A  
; CURRENT FILING DATE: 2001-11-21  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 261  
; TYPE: PRT  
; ORGANISM: human  
US-09-990-415A-2

Query Match	100.0%;	Score 1345;	DB 10;	Length 261;
Best local Similarity	100.0%;	Pred. No. 2.2e-141;		
Matches 261;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MESKMGELPLDINIOEPRWDQSTFLGRARHFTVTDPRLNLLSGAQLEASRNIVQNYRAG	60	
DB	1	MESKMGELPLDINIOEPRWDQSTFLGRARHFTVTDPRLNLLSGAQLEASRNIVQNYRAG	60	
QY	61	VVTPTGTEQDLRAKYVYDSAFHPDTGEKVVLIGRMSAQVPMNMTITGCMLEFYRKTPTV	120	
DB	61	VVTPTGTEQDLRAKYVYDSAFHPDTGEKVVLIGRMSAQVPMNMTITGCMLEFYRKTPTV	120	
QY	121	VFWQWVNSFNAINVYNSRSGDPTITVRQLGTAYVSATGAVATAIGKSLTKLPLVVG	180	
DB	121	VFWQWVNSFNAINVYNSRSGDPTITVRQLGTAYVSATGAVATAIGKSLTKLPLVVG	180	
QY	181	RFVFPAAVAANCINIPLMRQRELQVGI PVADAGQRLGYSVTAQKQIFQVVISRICMA	240	
DB	181	RFVFPAAVAANCINIPLMRQRELQVGI PVADAGQRLGYSVTAQKQIFQVVISRICMA	240	

QY 241 IPMAIPLIMDTLEKDFLK 261  
Db 241 IPMAIPLIMDTLEKDFLK 261

RESULT 2  
US-09-867-550-1344  
; Sequence 1344, Application US/09867550  
; Patent No. US20020082206A1  
; GENERAL INFORMATION:  
; APPLICANT: Leach, Martin D.  
; APPLICANT: Mehraban, Fuad,  
; APPLICANT: Conley, Pamela  
; APPLICANT: Law, Debbie  
; APPLICANT: Topper, James  
; TITLE OF INVENTION: No. US20020082206A1el Polynucleotides from Atherogenic Cells and  
; FILE REFERENCE: 21402-013 (Cura-313)  
; CURRENT APPLICATION NUMBER: US/09/867,550  
; PRIOR FILING DATE: 2001-09-20  
; PRIOR APPLICATION NUMBER: USSN 60/208,427  
; NUMBER OF SEQ ID NOS: 2125  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1344  
; LENGTH: 266  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-867-550-1344

Query Match 100.0%; Score 1345; DB 9; Length 266;  
Best Local Similarity 100.0%; Pred. No. 2.2e-141;  
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MESKNGELPLDINIOEPRWDQSTFLGRARHFTVTDPRNLISGAQLEASRNIVQNYRAG 60  
Db 1 MESKNGELPLDINIOEPRWDQSTFLGRARHFTVTDPRNLISGAQLEASRNIVQNYRAG 60  
QY 61 VVTPGITEQLMRAKYVYDSAFHPTDGEKVVLIGRMSAOVPMNMTTTGCMLEFYRKPTPV 120  
Db 61 VVTPGITEQLMRAKYVYDSAFHPTDGEKVVLIGRMSAOVPMNMTTTGCMLEFYRKPTPV 120  
QY 121 VFMQVWNSFNAIVNYSNRSGDPTITVRQLGTAYVSATGAVATAGLKSITKHLPLVVG 180  
Db 121 VFMQVWNSFNAIVNYSNRSGDPTITVRQLGTAYVSATGAVATAGLKSITKHLPLVVG 180  
QY 181 RFVPPAAVAAANCINIPLMRQRELOVGI PVADAGQRLGYSTAAKOGIFQVVISRICMA 240  
Db 181 RFVPPAAVAAANCINIPLMRQRELOVGI PVADAGQRLGYSTAAKOGIFQVVISRICMA 240  
QY 241 IPMAIPLIMDTLEKDFLK 261  
Db 241 IPMAIPLIMDTLEKDFLK 261

RESULT 3  
US-09-990-415A-8  
; Sequence 8, Application US/09990415A  
; Patent No. US20020165182A1  
; GENERAL INFORMATION:  
; APPLICANT: Pharmacia AB  
; TITLE OF INVENTION: Protein Cluster I  
; FILE REFERENCE: 00349  
; CURRENT APPLICATION NUMBER: US/09/990,415A  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 8  
; LENGTH: 322  
; TYPE: PRT  
; ORGANISM: human  
; FEATURE:  
; NAME/KEY: misc\_feature

; LOCATION: (25)..(25)  
; OTHER INFORMATION: Xaa=A,T,G or C  
US-09-990-415A-8

Query Match 81.0%; Score 1089; DB 10; Length 322;  
Best Local Similarity 78.5%; Pred. No. 1.1e-112;  
Matches 201; Conservative 28; Mismatches 27; Indels 0; Gaps 0;  
QY 6 GELPLDINIOEPRWDQSTFLGRARHFTVTDPRNLISGAQLEASRNIVQNYRAGVTPG 65  
Db 3 GELPPNINIKEPRWDQSTFLGRANHFTVTDPRNILLTNEQLESARKIVHDYRQGIYPFG 62  
QY 66 ITEDQLMRAKYVYDSAFHPTDGEKVVLIGRMSAOVPMNMTTTGCMLEFYRKPTPVVFWQW 125  
Db 63 LTENELMRAKYIYDSAFHPTDGEKVVLIGRMSAOVPMNMTTTGCMLEFYRKPTPVVFWQW 122  
QY 126 VNOSFNAIVNYSNRSGDPTITVRQLGTAYVSATGAVATAGLKSITKHLPLVGRFVVF 185  
Db 123 INOSFNAVNYNTRSGDAPLTVNELGTAYVSATGAVATAGLALTKHVSPLIGRFVVF 182  
QY 186 AAVAAANCINIPLMRQRELOVGI PVADAGQRLGYSTAAKOGIFQVVISRICMAIPAMA 245  
Db 183 AAVAAANCINIPLMRQRELOVGI PVTDENGNRLGESANAQAQAITQVVVSRIAMAAPGMA 242  
QY 246 IPPLIMDTLEKDFLK 261  
Db 243 IPPFIMNTLEKDFLK 258

RESULT 4  
US-10-014-338-2  
; Sequence 2, Application US/10014338  
; Publication No. US20030092614A1  
; GENERAL INFORMATION:  
; APPLICANT: Herach, et al.  
; TITLE OF INVENTION: ADP1-41, A NOVEL PROTEIN ISOLATED FROM BRAIN TISSUE HOMOGENATE A  
; FILE REFERENCE: 9195-077  
; CURRENT APPLICATION NUMBER: US/10/014,338  
; PRIOR FILING DATE: 2002-05-01  
; PRIOR APPLICATION NUMBER: 10/014,338  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 322  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-014-338-2

Query Match 81.0%; Score 1089; DB 15; Length 322;  
Best Local Similarity 78.5%; Pred. No. 1.1e-112;  
Matches 201; Conservative 28; Mismatches 27; Indels 0; Gaps 0;  
QY 6 GELPLDINIOEPRWDQSTFLGRARHFTVTDPRNLISGAQLEASRNIVQNYRAGVTPG 65  
Db 3 GELPPNINIKEPRWDQSTFLGRANHFTVTDPRNILLTNEQLESARKIVHDYRQGIYPFG 62  
QY 66 ITEDQLMRAKYVYDSAFHPTDGEKVVLIGRMSAOVPMNMTTTGCMLEFYRKPTPVVFWQW 125  
Db 63 LTENELMRAKYIYDSAFHPTDGEKVVLIGRMSAOVPMNMTTTGCMLEFYRKPTPVVFWQW 122  
QY 126 VNOSFNAIVNYSNRSGDPTITVRQLGTAYVSATGAVATAGLKSITKHLPLVGRFVVF 185  
Db 123 INOSFNAVNYNTRSGDAPLTVNELGTAYVSATGAVATAGLALTKHVSPLIGRFVVF 182  
QY 186 AAVAAANCINIPLMRQRELOVGI PVADAGQRLGYSTAAKOGIFQVVISRICMAIPAMA 245  
Db 183 AAVAAANCINIPLMRQRELOVGI PVTDENGNRLGESANAQAQAITQVVVSRIAMAAPGMA 242  
QY 246 IPPLIMDTLEKDFLK 261  
Db 243 IPPFIMNTLEKDFLK 258

```
RESULT 5
US-10-014-338-4
; Sequence 4, Application US/10014338
; Publication No. US2003009261A1
; GENERAL INFORMATION:
; APPLICANT: Heraeth, et al.
; TITLE OF INVENTION: ADPI-41, A NOVEL PROTEIN ISOLATED FROM BRAIN TISSUE HOMOGENATE AN
; FILE REFERENCE: USES THEREFOR
; FILE REFERENCE: 9195-077
; CURRENT APPLICATION NUMBER: US/10/014,338
; CURRENT FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: 10/014,338
; PRIOR FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 4
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-014-338-4
```

Query Match 65.1%; Score 876; DB 15; Length 261;  
Best Local Similarity 79.6%; Pred. No. 4.5e-89;  
Matches 156; Conservative 25; Mismatches 15; Indels 0; Gaps 0;

```
QY 6 GELPLDINIOEPRWDSTFLGRARHFTVTDPRNLLLSGAQLASRNIVQNYRAGVTPG 65
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 3 GELPPNINIKEPRWDSTFLGRARHFTVTDPRNLLLSGAQLASRNIVQNYRAGVTPG 62

QY 66 ITEDQIMRAKYVDSAFHPTDGEKVLIGRMSAQVPMNMTITGCMLEFYRKPTVVFWMQ 125
   :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 63 LTENELMRAKYVDSAFHPTDGEKMLIGRMSAQVPMNMTITGCMLEFYRKPTVVFWMQ 122

QY 126 VNOSFNALVNSNRSGDPTITVROLGTAVYSATGAVATAGLSLTGHLPLVGRFVPF 185
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 123 INQSFNAVNTNRSGDAPLTVNELGTAVYSVTGAVATAGLNALTKHVSPLIGRFVPF 182

QY 186 AAVAAANCINIPLMRQ 201
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 183 AAVAAANCINIPLMRQ 198
```

```
RESULT 6
US-09-990-415A-6
; Sequence 6, Application US/09990415A
; Patent No. US20020165182A1
; GENERAL INFORMATION:
; APPLICANT: Pharmacia AB
; TITLE OF INVENTION: Protein Cluster I
; FILE REFERENCE: 00349
; CURRENT APPLICATION NUMBER: US/09/990,415A
; CURRENT FILING DATE: 2001-11-21
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 6
; LENGTH: 322
; TYPE: PRT
; ORGANISM: human
US-09-990-415A-6
```

Query Match 56.0%; Score 753; DB 10; Length 322;  
Best Local Similarity 59.0%; Pred. No. 3.2e-75;  
Matches 147; Conservative 37; Mismatches 65; Indels 0; Gaps 0;

```
QY 13 NIOEPRWDSTFLGRARHFTVTDPRNLLLSGAQLASRNIVQNYRAGVTPGITEDQIM 72
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 9 NIDAPRWDQRTFLGRVKHFLNITDPRTVFVSERELDWAQVVEKSRMGVPPGTQVEQLL 68

QY 73 RAKYVDSAFHPTDGEKVLIGRMSAQVPMNMTITGCMLEFYRKPTVVFWMQVNSFNA 132
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 69 YAKKLYDSAFHPTDGEKMNVIGRMSFQLPGMIIITGFMLOFYRTMPAVIFWQVNSFNA 128
```

```
QY 133 IVNYSNRSGDPTITVROLGTAVYSATGAVATAGLSLTGHLPLVGRFVPFAVAAN 192
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 129 LVNTYTRNAASPTSVROMALSYFTATTAVATAGVMMLTKKAPPLVGRVFPFAVAAN 188

QY 193 CINIPLMRQRELOVGIPIVADGAGQRLGYSVTAAKOGIFQVVISRICMAIPMAIPPLIMD 252
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 189 CVNIPMRQRELKIGICVKDRNENEIGHSRRAAIGITQVVISRITMSAPGMILLPVIME 248

QY 253 TLEKDFLK 261
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 249 RLEKLHFMQ 257
```

```
RESULT 7
US-10-094-749-2165
; Sequence 2165, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOTYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 08435/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2165
; LENGTH: 186
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2165
```

Query Match 32.4%; Score 436; DB 12; Length 186;  
Best Local Similarity 56.5%; Pred. No. 3.3e-40;  
Matches 83; Conservative 22; Mismatches 42; Indels 0; Gaps 0;

```
QY 3 SKMGELPLDINIOEPRWDSTFLGRARHFTVTDPRNLLLSGAQLASRNIVQNYRAGV 62
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 17 SKMEADLSGFNIDAPRWDQRTFLGRVKHFLNITDPRTVFVSERELDWAQVVEKSRMGV 76

QY 63 TPGITEDQIMRAKYVDSAFHPTDGEKVLIGRMSAQVPMNMTITGCMLEFYRKPTVVF 122
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 77 PPGTQVEQLLYAKKLYDSAFHPTDGEKMNVIGRMSFQLPGMIIITGFMLOFYRTMPAVIF 136

QY 123 WQVNSFNALVNSNRSGDPTITVRQ 149
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 137 WQVNSFNALVNTNRNAASPTSVRK 163
```

```
RESULT 8
US-09-990-415A-4
; Sequence 4, Application US/09990415A
```

```

; Patent No. US20020165182A1
; GENERAL INFORMATION:
; APPLICANT: Pharmacia AB
; TITLE OF INVENTION: Protein cluster I
; FILE REFERENCE: 00349
; CURRENT APPLICATION NUMBER: US/09/990,415A
; CURRENT FILING DATE: 2001-11-21
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 77
; TYPE: PRT
; ORGANISM: human
US-09-990-415A-4

```

Query Match	27.4%	Score 368.5;	DB 10;	Length 77;
Best Local Similarity	57.5%;	Pred. No. 3.1e-33;		
Matches 77; Conservative	0;	Mismatches 0;	Indels 57;	Gaps 1.

```
QY      1 MESKMGELPLDINIQEPRWDSTFLGRARHFTVTDPRNLLSGAQLASRNIYQNTRAG 60
        |||||
Db       1 MESKMGELPLDINIQEPRWDQSTFLGRARHFTVTDPRNLLSQAQLASRNIYN----- 56
QY      61 VVTPGITEDQLMRAKYVYDSAFHPDTGEKVVLIGRMSAQVPMNMTITGCMLTFYRKPTV 120
        |||||
Db       57 -----YRKPTV 63
QY      121 VFQWVNQSFNALIV 134
        |||||
Db       64 VFQWVNQSFNALIV 77
```

RESULT 9  
US-10-264-237-2468

```

; Sequence 2468, Application US/10264237
; Publication No. US20040009491A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA131P1
; CURRENT APPLICATION NUMBER: US/10/264,237
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/16450
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/205,515
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 2876
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 2468
;
; LENGTH: 127
;
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-264-237-2468

```

Query Match	23.2%;	Score 312;	DB 12;	Length 127;
Best Local Similarity	100.0%;	Pred. No. 1.3e-26;		
Matches 63; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0

[illegible]

RESULT 10  
US-10-415-378-1  
; Sequence 1, Application US/10415378  
; Publication No. US20040014945A1  
; GENERAL INFORMATION:  
; APPLICANT: INCYTE CORPORATION; TANG, Y. Tom

```

; APPLICANT: YUE, Henry; NGUYEN, Daniel B.;
; APPLICANT: HAFALIA, April J.A.; ELLIOTT, Vicki S.;
; APPLICANT: LU, Yan; CHAWLA, Narender K.;
; APPLICANT: YAO, Monique G.; BAUGHN, Mariah R.;
; APPLICANT: GANDHI, Ameena R.; DING, Li;
; APPLICANT: SANJANWALA, Madhusudan M.; RAMKUMAR, Jayalaxmi;
; APPLICANT: ARVIZU, Chandra S.; GIERZEN, Kimberly J.;
; APPLICANT: LAL, Preeti G.; AZIMZAI, Yalda;
; APPLICANT: KHAN, Farrah A.; THANGAVELU, Kavitha;
; APPLICANT: THORNTON, Michael B.; LU, Dyung Aina M.;
; APPLICANT: TRIBOUTEY, Catherine M.; WARREN, Bridget A.;
; APPLICANT: ISON, H. Craig; DAS, Debopriya;
; APPLICANT: RAUMANN, Brigitte E.; POLICKY, Jennifer L.;
; APPLICANT: KEARNEY, Liam
; TITLE OF INVENTION: TRANSPORTERS AND ION CHANNELS
; FILE REFERENCE: PI-0270 USN
; CURRENT APPLICATION NUMBER: US/10/415,378
; PRIOR FILING DATE: 2003-05-07
; PRIOR APPLICATION NUMBER: PCT/US01/46055
; PRIOR FILING DATE: 2001-10-27
; PRIOR APPLICATION NUMBER: US 60/250,790
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: US 60/252,232
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/249,661
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: US 60/247,673
; PRIOR FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/245,904
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: US 60/243,989
; PRIOR FILING DATE: 2000-10-27
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PERL Program
; SEQ ID NO 1
; LENGTH: 337
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20040014945A1 1626101CD1
US-10-415-378-1

```

Query Match	10.7%;	Score 143.5;	DB 12;	Length 337;
Best Local Similarity	22.1%;	Pred. No. 3.4e-07;		
Matches	57;	Conservative	56;	Mismatches 128;
				Indels 17;
				Gaps 6;

[illegible]

```

RESULT 11
US-10-106-698-6940
; Sequence 6940, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:

```





FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0463 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
TELEX: <Unknown>  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 344 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 755185  
SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-10-306-133-3

Query Match 6.7%; Score 89.5; DB 15; Length 344;  
Best Local Similarity 24.8%; Pred. No. 0.37;  
Matches 52; Conservative 28; Mismatches 95; Indels 35; Gaps 10;

QY 2 ESKMGELPLDINIQEPRW-DOSTFLGRARHFTVDPRLNLLSGAQLASRNI--VQNYR 58  
DB 51 ESATLRCTIDNRYTRVAVMLNRSTILYAGNDKXCL-DPRVLLSNTQYQYSEIQNDVYD 109  
QY 59 AGVTPGITEPDQWRAKYVYDSAFHPDTEKVVLIGRMSAQV-----PMNMTIT 107  
DB 110 EGPYTCVQTDN-----HPKT-SRVHLIVQVSPKIVEISSDISINEGNISLT 156  
QY 108 GCMLTFYRKPTPTVFWQVWNSFNALVYNSRSGDTPITVRQLGTAVYSATGAVATALG 167  
DB 157 -CIAT-GRPEPTVT-WRHISPKAVGFVSEDEVLEIQGITRQSGEYECASNDVAAPVVR 213  
QY 168 LKSLTKHLPPLV----GRFVPFAVAALANC 193  
DB 214 RVNVTVNYPPIYSEAKGTGVPVGQKGTLOC 243

## RESULT 15

US-09-978-295A-523  
Sequence 523, Application US/09978295A  
Patent No. US2002015606A1  
GENERAL INFORMATION:  
APPLICANT: Ashkenazi, Avi  
APPLICANT: Baker Kevin P.  
APPLICANT: Botstein, David  
APPLICANT: Desnoyers, Luc  
APPLICANT: Eaton, Dan  
APPLICANT: Ferrara, Napoleone  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Fong, Sherman  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerber, Hanspeter  
APPLICANT: Gerritsen, Mary E.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Grimaldi, J. Christopher  
APPLICANT: Gurney, Austin L.  
APPLICANT: Hillan, Kenneth J.  
APPLICANT: Kljavin, Ivar J.  
APPLICANT: Kuo, Sophia S.  
APPLICANT: Napier, Mary A.  
APPLICANT: Pan, James  
APPLICANT: Paoni, Nicholas F.  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Shelton, David L.  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Williams, P. Mickey

APPLICANT: Wood, William I.  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
FILE REFERENCE: P2630P1C11  
CURRENT APPLICATION NUMBER: US/09/978,295A  
PRIORITY FILING DATE: 2001-10-15  
PRIOR APPLICATION NUMBER: 09/918585  
PRIOR FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: 60/062250  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/064249  
PRIOR FILING DATE: 1997-11-03  
PRIOR APPLICATION NUMBER: 60/065311  
PRIOR FILING DATE: 1997-11-13  
PRIOR APPLICATION NUMBER: 60/066364  
PRIOR FILING DATE: 1997-11-21  
PRIOR APPLICATION NUMBER: 60/077450  
PRIOR FILING DATE: 1998-03-10  
PRIOR APPLICATION NUMBER: 60/077632  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077641  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077649  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077791  
PRIOR FILING DATE: 1998-03-12  
PRIOR APPLICATION NUMBER: 60/078004  
PRIOR FILING DATE: 1998-03-13  
PRIOR APPLICATION NUMBER: 60/078886  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078936  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078910  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078939  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/079294  
PRIOR FILING DATE: 1998-03-25  
PRIOR APPLICATION NUMBER: 60/079656  
PRIOR FILING DATE: 1998-03-26  
PRIOR APPLICATION NUMBER: 60/079664  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079689  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079728  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079786  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079920  
PRIOR FILING DATE: 1998-03-30  
PRIOR APPLICATION NUMBER: 60/079923  
PRIOR FILING DATE: 1998-03-30  
PRIOR APPLICATION NUMBER: 60/080105  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080107  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080165  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080194  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080327  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080328  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080333  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080334  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/081070  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081049

PRIOR FILING DATE: 1998-04-08	PRIOR APPLICATION NUMBER: 60/081071
PRIOR FILING DATE: 1998-04-08	PRIOR APPLICATION NUMBER: 60/081195
PRIOR FILING DATE: 1998-04-08	PRIOR APPLICATION NUMBER: 60/081203
PRIOR FILING DATE: 1998-04-09	PRIOR APPLICATION NUMBER: 60/081229
PRIOR FILING DATE: 1998-04-09	PRIOR APPLICATION NUMBER: 60/081955
PRIOR FILING DATE: 1998-04-15	PRIOR APPLICATION NUMBER: 60/081819
PRIOR FILING DATE: 1998-04-15	PRIOR APPLICATION NUMBER: 60/081819
PRIOR FILING DATE: 1998-04-15	PRIOR APPLICATION NUMBER: 60/081952
PRIOR FILING DATE: 1998-04-15	PRIOR APPLICATION NUMBER: 60/081838
PRIOR FILING DATE: 1998-04-15	PRIOR APPLICATION NUMBER: 60/082568
PRIOR FILING DATE: 1998-04-21	PRIOR APPLICATION NUMBER: 60/082569
PRIOR FILING DATE: 1998-04-21	PRIOR APPLICATION NUMBER: 60/082704
PRIOR FILING DATE: 1998-04-22	PRIOR APPLICATION NUMBER: 60/082804
PRIOR FILING DATE: 1998-04-22	PRIOR APPLICATION NUMBER: 60/082700
PRIOR FILING DATE: 1998-04-22	PRIOR APPLICATION NUMBER: 60/082797
PRIOR FILING DATE: 1998-04-22	PRIOR APPLICATION NUMBER: 60/082796
PRIOR FILING DATE: 1998-04-23	PRIOR APPLICATION NUMBER: 60/083336
PRIOR FILING DATE: 1998-04-27	PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28	PRIOR APPLICATION NUMBER: 60/083392
PRIOR FILING DATE: 1998-04-29	PRIOR APPLICATION NUMBER: 60/083495
PRIOR FILING DATE: 1998-04-29	PRIOR APPLICATION NUMBER: 60/083496
PRIOR FILING DATE: 1998-04-29	PRIOR APPLICATION NUMBER: 60/083499
PRIOR FILING DATE: 1998-04-29	PRIOR APPLICATION NUMBER: 60/083545
PRIOR FILING DATE: 1998-04-29	PRIOR APPLICATION NUMBER: 60/083554
PRIOR FILING DATE: 1998-04-29	PRIOR APPLICATION NUMBER: 60/083558
PRIOR FILING DATE: 1998-04-29	PRIOR APPLICATION NUMBER: 60/083559
PRIOR FILING DATE: 1998-04-29	PRIOR APPLICATION NUMBER: 60/083500
PRIOR FILING DATE: 1998-04-29	PRIOR APPLICATION NUMBER: 60/083742
PRIOR FILING DATE: 1998-04-30	PRIOR APPLICATION NUMBER: 60/084366
PRIOR FILING DATE: 1998-05-05	PRIOR APPLICATION NUMBER: 60/084414
PRIOR FILING DATE: 1998-05-06	PRIOR APPLICATION NUMBER: 60/084411
PRIOR FILING DATE: 1998-05-06	PRIOR APPLICATION NUMBER: 60/084637
PRIOR FILING DATE: 1998-05-07	PRIOR APPLICATION NUMBER: 60/084639
PRIOR FILING DATE: 1998-05-07	PRIOR APPLICATION NUMBER: 60/084640
PRIOR FILING DATE: 1998-05-07	PRIOR APPLICATION NUMBER: 60/084598
PRIOR FILING DATE: 1998-05-07	PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07	

```

1  PRIOR APPLICATION NUMBER: 60/084627
2  PRIOR FILING DATE: 1998-05-07
3  PRIOR APPLICATION NUMBER: 60/084643
4  PRIOR FILING DATE: 1998-05-07
5  PRIOR APPLICATION NUMBER: 60/085339
6  PRIOR FILING DATE: 1998-05-13
7  PRIOR APPLICATION NUMBER: 60/085338
8  PRIOR FILING DATE: 1998-05-13
9  PRIOR APPLICATION NUMBER: 60/085323
10 PRIOR FILING DATE: 1998-05-13
11 PRIOR APPLICATION NUMBER: 60/085582
12 PRIOR FILING DATE: 1998-05-15
13 PRIOR APPLICATION NUMBER: 60/085700
14 PRIOR FILING DATE: 1998-05-15
15 PRIOR APPLICATION NUMBER: 60/085689
16 PRIOR FILING DATE: 1998-05-15
17 PRIOR APPLICATION NUMBER: 60/085579
18 PRIOR FILING DATE: 1998-05-15
19 PRIOR APPLICATION NUMBER: 60/085580
20 PRIOR FILING DATE: 1998-05-15
21 PRIOR APPLICATION NUMBER: 60/085573
22 PRIOR FILING DATE: 1998-05-15
23 PRIOR APPLICATION NUMBER: 60/085704
24 PRIOR FILING DATE: 1998-05-15
25 PRIOR APPLICATION NUMBER: 60/085697

```

Query Match	6.6%;	Score 88.5;	DB 10;	Length 344;
Best Local Similarity	24.8%;	Pred. No. 0.48;		
Matches	52;	Conservative	27;	Mismatches 96;
				Indels 35;
				Gaps 10;

```

QY      2 ESKMGELPLDINIDEPRW-DOSTFLGRARHFFVTVPDRNLLSGAOLEASRNI--VQNYR  58
      ||      :      :      :      :      :      :      :      :      :
Db      51 ESATLRCTIDNRVTRVAAWLNRSSTILYAGNDKWCL-DPRVLLSNTQTQYSIEIQNVVDYD  109

QY      59 AGVTPGITEEDQLWRAKYVYDSAFHPDTEKVLIGRMSAQV-----PMNMTIT  107
      ||      :      :      :      :      :      :      :      :      :
Db      110 EGPYTCVQTDN-----HPKT-SRVHLIVQVSPKIVEISDISINEGNNISLT  156

QY      108 GCMLTFYRKPTFVVEFWQWVNSFNALVYNSNRSGDPTITVRQLGFAVVSATTGAVATALG  167
      ||      :      :      :      :      :      :      :      :      :
Db      157 -CIAT-GRPEPTVT-WRHISPKAVGVSEDEYLEIGCITREQSGDYECASANDVAAPVR  213

QY      168 LKSLTKHLPLV----GRFVPFAAVALAANC  193
      :      :      :      :      :      :      :      :      :
Db      214 RUKVTVNYPPIYSBAKGTGVPGQKGTLQC  243

```

Search completed: February 3, 2004, 09:16:01  
Job time : 35 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 3, 2004, 09:11:48 ; Search time 41 Seconds  
(without alignments)  
1010.430 Million cell updates/sec

Title: US-09-990-415A-2  
Perfect score: 1345  
Sequence: 1 MESKMGEPLDINIQEPRWD.....PAMAIPLIMDTLEKKDFLK 261

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues  
Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_19Jun03.\*  
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*  
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1345	100.0	261	23	ABB76445	Human protein clus
2	1345	100.0	266	23	ABP64302	Human ORF672. Hom
3	1345	100.0	312	23	ABB06151	Human NS protein 8
4	1345	100.0	325	22	AAAM93760	Human polypeptide,
5	1345	100.0	374	22	AAAM39967	Human polypeptide
6	1345	100.0	374	22	AAAB60083	Human transport pr
7	1269	94.3	251	21	AAAB41589	Human ORFX ORF1353
8	1255	93.3	397	22	ABG20175	Novel human diagno
9	1089	81.0	322	22	AAAM40413	Human polypeptide

10	1089	81.0	322	23	ABB79572	Alzheimer's diseas
11	1089	81.0	322	23	ABG68039	Human Alzheimer's
12	1089	81.0	322	23	ABB76448	Human protein clus
13	1089	81.0	322	23	AAE13276	Human transporters
14	1089	81.0	332	21	AAAB41585	Human ORFX ORF1349
15	1089	81.0	351	22	AAAM42198	Human polypeptide
16	1082	80.4	322	22	AAAG62389	Rat tricarboxylate
17	977	72.6	325	23	AAU91141	Human secreted pro
18	975	72.5	230	23	AAU91139	Human secreted pro
19	975	72.5	251	23	AAU91140	Human secreted pro
20	876	65.1	261	23	ABB79573	Alzheimer's diseas
21	876	65.1	261	23	ABG68040	Amino acid sequenc
22	851	63.3	242	22	AAAB93934	Human protein sequ
23	754	56.1	321	22	ABB65884	Drosophila melanog
24	753	56.0	322	23	ABB76447	Human protein clus
25	750	55.8	420	22	AAAM40491	Human polypeptide
26	749	55.7	322	22	AAAM38705	Human polypeptide
27	749	55.7	322	23	AAE21177	Human TRICH-21 pro
28	726	54.0	236	22	AAAB95218	Human protein sequ
29	636.5	47.3	195	22	AAU32503	Novel human secret
30	602	44.8	174	22	AAU19596	Human diagnostic a
31	593.5	44.1	351	22	ABB62710	Drosophila melanog
32	545	40.5	132	22	ABG20174	Novel human diagno
33	545	40.5	203	23	ABB97807	Human secretory po
34	503	37.4	112	23	ABB79571	Alzheimer's diseas
35	480.5	35.7	340	23	AAE22908	Human transporter
36	368.5	27.4	77	23	ABB76446	Human protein clus
37	312	23.2	127	23	ABB90092	Human polypeptide
38	301	22.4	138	22	AAAM41753	Human secreted pro
39	258	19.2	98	21	AAAG03234	Human transporter
40	143.5	10.7	337	23	ABG61531	Human protein sequ
41	143.5	10.7	352	22	AAAM25384	Human colon cancer
42	143	10.6	54	22	AAAG76166	Human zinc-finger
43	139	10.3	305	23	ABB06187	Human secreted pro
44	117	8.7	79	21	AAAG02779	Angiotensin conver
45	96	7.1	775	22	AAU02904	

ALIGNMENTS

RESULT 1  
ABB76445 ABB76445 standard; Protein; 261 AA.  
XX ABB76445;  
AC  
XX  
DT 02-SEP-2002 (first entry)  
XX  
DE Human protein cluster I polypeptide.  
XX  
KW Protein cluster I; human; metabolic disorder; obesity; diabetes;  
KW antidiabetic; diagnosis; therapy.  
XX  
OS Homo sapiens.  
XX  
PN WO200242324-A1.  
XX  
PD 30-MAY-2002.  
XX  
PF 22-NOV-2001; 2001WO-SE02581.  
XX  
PR 24-NOV-2000; 2000SE-0004325.  
XX  
PA (PHAA ) PHARMACIA AB.  
XX  
PI Atterband A;  
XX  
XX  
DR WPI; 2002-500277/53.  
DR N-PSDB; ABN83754.  
XX  
PT Novel nucleic acid molecule encoding Protein Cluster I, useful in the  
PT diagnosis of metabolic diseases, such as obesity and diabetes, and in



PT the identification of agents useful in the treatment of the diseases -  
XX  
XX  
PS Claim 3; Page 24-25; 34pp; English.  
XX  
CC The present invention relates to the identification of a human gene  
CC family (see ABN83754-57) encoding a group of polypeptides (see  
CC ABB76445-48) referred to as Protein Cluster I. This family of  
CC homologous proteins was identified by an 'all-versus-all' BLAST  
CC procedure using all Caenorhabditis elegans proteins in the  
CC Wormpep20 database release. Proteins of unknown function were  
CC compared to the Drosophila melanogaster Flybase database, and  
CC non-annotated protein clusters, conserved in both C. elegans and D.  
CC melanogaster, were used in a BLAST procedure against the Celera  
CC Human Genome Database, and Protein Cluster I proteins of unknown  
CC function were selected for study. The human part of Protein  
CC Cluster I comprises polypeptides encoded by 3 genes. The partial  
CC gene sequence encoding the present polypeptide is expressed  
CC primarily in the nervous system and digestive system. The claimed  
CC Protein Cluster I nucleic acid molecules and proteins are proposed  
CC to be useful for differential identification of the tissue(s) or  
CC cell type(s) present in a biological sample, for diagnosis of  
CC diseases and disorders, including metabolic disorders and immune  
CC disorders, especially obesity and diabetes, and for identifying  
CC agents useful in the treatment of such diseases.  
XX

XX Sequence 261 AA;

Query Match 100.0%; Score 1345; DB 23; Length 261;  
Best Local Similarity 100.0%; Pred. No. 2.1e-144;  
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MESKMGELPLDINIQEPRWDQSTFLGRARHFFVTDPRLNLLSGAQLASRNIVQNYRAG 60  
DB 1 MESKMGELPLDINIQEPRWDQSTFLGRARHFFVTDPRLNLLSGAQLASRNIVQNYRAG 60  
QY 61 VTPGITEDQLMRAKYVYDSAFHPDTEKVVLLIGRMSAQVPMNMTITGCMLETFYRKPTV 120  
DB 61 VTPGITEDQLMRAKYVYDSAFHPDTEKVVLLIGRMSAQVPMNMTITGCMLETFYRKPTV 120  
QY 121 VFWQWVNSFNAINVYSNRSGDPTITVRQLGTAYVSATTGAVATAGLKSITKHLPLVVG 180  
DB 121 VFWQWVNSFNAINVYSNRSGDPTITVRQLGTAYVSATTGAVATAGLKSITKHLPLVVG 180  
QY 181 RFVPPFAAANAANCINIPLMRQRELQVGIIPVADEAGQRLGYSVTAAGQIFQVVISRICMA 240  
DB 181 RFVPPFAAANAANCINIPLMRQRELQVGIIPVADEAGQRLGYSVTAAGQIFQVVISRICMA 240  
QY 241 IPAMAIPPLIMDTLEKDFLK 261  
DB 241 IPAMAIPPLIMDTLEKDFLK 261

RESULT 2  
ABP64302  
ID ABP64302 standard; Protein; 266 AA.

XX  
AC ABP64302;

DT 04-NOV-2002 (first entry)

DE Human ORF672.

KW Cytostatic; Cardiant; Anti-allergic; Immunosuppressive; Vulnerary;  
KW Antiinflammatory; gene therapy; human; ORFX; atherogenic; platelet;  
KW human umbilical vein endothelial cell; HUVEC; atherosclerotic plaque;  
KW cancer; cardiovascular disease; allergy; autoimmune disease;  
KW wound healing; blood coagulation disorder; inflammatory disorder.

OS Homo sapiens.  
XX  
XX US2002082206-A1.  
PN  
XX 27-JUN-2002.

XX 30-MAY-2001; 2001US-0867550.  
PF  
XX  
XX 30-MAY-2000; 2000US-208427P.  
PR  
XX  
PA (LEAC/) LEACH M D.  
PA (MEHR/) MEHRABAN F.  
PA (CONL/) CONLEY P B.  
PA (TOPP/) TOPPER J N.  
PA (LAWD/) LAW D.  
XX  
PI Leach MD, Mehraban F, Conley PB, Topper JN, Law D;  
XX  
DR WPI; 2002-626554/67;  
DR N-PSDB; ABQ98865.  
XX  
PT New polypeptide designated ORFX are present in human atherogenic cells  
PT and are useful to prevent and treat ORFX-associated disorders including  
PT cancer, allergy, wound healing or autoimmune, cardiovascular or  
PT inflammatory disease -  
XX  
XX Claim 10; SEQ ID 1344; 78pp; English.

XX The present invention relates to novel human ORFX polypeptides and their  
XX coding sequences (ABP63631-ABP64681 and ABQ98194-ABQ99267). The sequences  
XX were discovered in human atherogenic cells, in particular in platelets  
XX and human umbilical vein endothelial cells (HUVEC) and are expressed in  
XX many other tissues as well. Atherogenic cells are cells which have the  
XX potential to develop atherosclerotic plaques. The ORFX polypeptides and  
XX nucleic acids are useful for treating or preventing a pathological  
XX condition associated with an ORFX-associated disorder, e.g. cancer,  
XX cardiovascular disease, allergy, autoimmune disease, wound healing, blood  
XX coagulation disorders or inflammatory disorders.  
XX Note: The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from the  
XX USPTO web site at seqdata.uspto.gov/sequence.html?docid=20020082206.

XX Sequence 266 AA;

Query Match 100.0%; Score 1345; DB 23; Length 266;  
Best Local Similarity 100.0%; Pred. No. 2.2e-144;  
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MESKMGELPLDINIQEPRWDQSTFLGRARHFFVTDPRLNLLSGAQLASRNIVQNYRAG 60  
DB 1 MESKMGELPLDINIQEPRWDQSTFLGRARHFFVTDPRLNLLSGAQLASRNIVQNYRAG 60  
QY 61 VTPGITEDQLMRAKYVYDSAFHPDTEKVVLLIGRMSAQVPMNMTITGCMLETFYRKPTV 120  
DB 61 VTPGITEDQLMRAKYVYDSAFHPDTEKVVLLIGRMSAQVPMNMTITGCMLETFYRKPTV 120  
QY 121 VFWQWVNSFNAINVYSNRSGDPTITVRQLGTAYVSATTGAVATAGLKSITKHLPLVVG 180  
DB 121 VFWQWVNSFNAINVYSNRSGDPTITVRQLGTAYVSATTGAVATAGLKSITKHLPLVVG 180  
QY 181 RFVPPFAAANAANCINIPLMRQRELQVGIIPVADEAGQRLGYSVTAAGQIFQVVISRICMA 240  
DB 181 RFVPPFAAANAANCINIPLMRQRELQVGIIPVADEAGQRLGYSVTAAGQIFQVVISRICMA 240  
QY 241 IPAMAIPPLIMDTLEKDFLK 261  
DB 241 IPAMAIPPLIMDTLEKDFLK 261

RESULT 3  
ABB06151  
ID ABB06151 standard; Protein; 312 AA.

XX  
XX ABB06151;  
AC  
XX  
XX 10-MAY-2002 (first entry)  
DT  
XX  
XX Human NS protein sequence SEQ ID NO:243.

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;  
KW antirheumatic; antiarthritic; antipsoriatic; ophthalmological; anti-HIV;  
KW vasotropic; antiarteriosclerotic; antiinflammatory; dermatological;  
KW anorectic; muscular; antiinfertility; cardiovascular; anticoagulant;  
KW antifibrinolytic; hypotension; antiasthmatic; immunomodulator; cardiant;  
KW anticonvulsant; antidiabetic; tranquilliser; antidepressant; neuroleptic;  
KW gastrointestinal; virucide; antitumor; cerebroprotective; nootropic;  
KW contraceptive; vaccine; gene therapy; cancer; osteoporosis; dystonia;  
KW endometriosis; degenerative disease; multiple sclerosis; psoriasis;  
KW rheumatoid arthritis; cataract; restenosis; atherosclerosis; glaucoma;  
KW inflammation; skin disorder; obesity; muscular dystrophy; AIDS;  
KW infertility; cardiovascular disease; coagulation disease; hypertension;  
KW ischaemia; asthma; immune disease; epilepsy; angina; neurodegeneration;  
KW diabetes; anxiety; depression; schizophrenia; viral disease; stroke;  
KW gastric ulcer; Alzheimer's disease.

XX Homo sapiens.

OS WO200206315-A2.

PN 24-JAN-2002.

PD 17-JUL-2001; 2001WO-IL00653.

PF 18-JUL-2000; 2000IL-0137345.

PR 15-DEC-2000; 2000IL-0140354.

XX (COMP-) COMPUGEN LTD.

XX Mintz L, Freilich S, Bernstein J;

XX WPI; 2002-155037/20.

DR N-PSDB; ABL39805.

PT One hundred and twenty eight novel nucleic acid sequences, useful for  
PT treating and diagnosing e.g. cancer, asthma and Alzheimer's -

PS Claim 6; Page 278-279; 290pp; English.

XX ABL39691 to ABL39818 represent novel human nucleic acid sequences  
CC encoding the proteins given in ABB06037 to ABB06164. The novel sequences  
CC (NS) can have cytostatic, osteopathic, gynaecological, neuroprotective,  
CC antirheumatic, antiarthritic, antipsoriatic, ophthalmological, virucide,  
CC vasotropic, antiarteriosclerotic, antiinflammatory, dermatological,  
CC anorectic, muscular, anti-HIV, antiinfertility, cardiovascular,  
CC anticoagulant, antifibrinolytic, hypotension, antiasthmatic, cardiant,  
CC immunomodulator, anticonvulsant, antidiabetic, tranquilliser, antitumor,  
CC antidepressant, gastrointestinal, neuroleptic, cerebroprotective,  
CC nootropic and contraceptive activities. The NS can be used in vaccines,  
CC gene therapy and antisense therapy. Nucleic acids, expression vectors and  
CC antibodies from the present invention can be used for treating and  
CC diagnosing e.g. cancer, osteoporosis, endometriosis, degenerative  
CC diseases, dystonia, multiple sclerosis, rheumatoid arthritis, psoriasis,  
CC cataracts, restenosis, atherosclerosis, inflammation, skin disorders,  
CC glaucoma, obesity, muscular dystrophy, AIDS, infertility, cardiovascular  
CC disease, coagulation disease, ischaemia, hypertension, diabetes, immune  
CC disease, epilepsy, angina, neurodegeneration, diabetes, anxiety,  
CC depression, schizophrenia, viral disease, gastric ulcers, stroke,  
CC Alzheimer's disease and as a contraceptive.

XX Sequence 312 AA;

Query Match 100.0%; Score 1345; DB 23; Length 312;  
Best Local Similarity 100.0%; Pred. No. 2.8e-144;  
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MESKMGEPLDINIQEPRWDQSTFLGRARHFTVTDPRLNLLSGAOLEASRNIVQNYRAG 60  
Db 1 MESKMGEPLDINIQEPRWDQSTFLGRARHFTVTDPRLNLLSGAOLEASRNIVQNYRAG 60  
OY 61 VVTGITEQLWRAKYVYDSAFHPDTGKRVLLIGRMSAQVPMNMTITGCMLEFYRKPTV 120

Db 61 VVTGITEQLWRAKYVYDSAFHPDTGKRVLLIGRMSAQVPMNMTITGCMLEFYRKPTV 120  
OY 121 VFWQWVNSFNAINVNSRSGDPTITVRQLGTAYVSATTGAVATAGLSLTKHLPLVVG 180  
Db 121 VFWQWVNSFNAINVNSRSGDPTITVRQLGTAYVSATTGAVATAGLSLTKHLPLVVG 180  
OY 181 RFVPEFAVAANCINIPLMRQRELVGIPVADAGQRLGYSTAAKQGIQVVISRICMA 240  
Db 181 RFVPEFAVAANCINIPLMRQRELVGIPVADAGQRLGYSTAAKQGIQVVISRICMA 240  
OY 241 IPMAIPLIMDTLEKDFLK 261  
Db 241 IPMAIPLIMDTLEKDFLK 261

RESULT 4

AAM93760 AAM93760 standard; Protein; 325 AA.

AC AAM93760;

DT 06-NOV-2001 (first entry)

DE Human polypeptide, SEQ ID NO: 3752.

KW Human; full length cDNA; cDNA synthesis; oligo-capping.

OS Homo sapiens.

PN EP1130094-A2.

PD 05-SEP-2001.

PF 07-JUL-2000; 2000EP-0114089.

PR 08-JUL-1999; 99JP-0194486.

PR 11-JAN-2000; 2000JP-0118774.

PR 02-MAY-2000; 2000JP-0183765.

XX (HELI-) HELIX RES INST.

XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;

PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

XX WPI; 2001-524255/58.

DR N-PSDB; AAK94713.

PT 830 Primers useful for synthesizing full length cDNA clones and their  
PT use in genetic manipulation -

PS Claim 8; SEQ ID NO 3752; 1380pp + sequence listing; English.

XX The invention relates to primers for synthesizing full length cDNA  
CC clones. 830 cDNA molecules encoding a human protein have been  
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA  
CC molecules have been determined. Primers for synthesizing the full length  
CC cDNA are useful for clarifying the function of the protein encoded by  
CC the cDNA. The full length clones were obtained by construction of full  
CC length enriched cDNA libraries that were synthesised by the oligo-capping  
CC method. The primers enable the production of the full length cDNA easily  
CC without any special methods. The present sequence is a polypeptide  
CC encoded by a full length human cDNA of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in CD-ROM format directly from EPO.

XX Sequence 325 AA;

Query Match 100.0%; Score 1345; DB 22; Length 325;  
Best Local Similarity 100.0%; Pred. No. 3e-144;  
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MESKMGEPLDINIQEPRWDQSTFLGRARHFTVTDPRLNLLSGAOLEASRNIVQNYRAG 60

Db 1 MESKMGELPLDINIOEPRWDSTFLGRARHFTVTDPRLNLISGAQLEASRNIVQNYRAG 60  
QY 61 VVTPGITEDQLMRAKYVYDSAFHPDTEGKVVLIGRMSAQVPMNMTTTGCMLEFYRKTPTV 120  
Db 61 VVTPGITEDQLMRAKYVYDSAFHPDTEGKVVLIGRMSAQVPMNMTTTGCMLEFYRKTPTV 120  
QY 121 VFWQWVQNSFNAIVNYSNRSGDPTITVRQLGTAYVSATTGAVATAGLKSCLKPLPLVG 180  
Db 121 VFWQWVQNSFNAIVNYSNRSGDPTITVRQLGTAYVSATTGAVATAGLKSCLKPLPLVG 180  
QY 181 RFVFPFAVAANANCINIPLMRQRELOVGIPVADEAGQRLGYSVTAAKOGIFQVVISRICMA 240  
Db 181 RFVFPFAVAANANCINIPLMRQRELOVGIPVADEAGQRLGYSVTAAKOGIFQVVISRICMA 240  
QY 241 IPAMAIPLIMDTLEKDFLK 261  
Db 241 IPAMAIPLIMDTLEKDFLK 261  
RESULT 5  
AAM39967  
ID AAM39967 standard; Protein; 374 AA.  
AC AAM39967;  
XX 22-OCT-2001 (first entry)  
DT Human polypeptide SEQ ID NO 3112.  
DE  
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia.  
XX  
OS Homo sapiens.  
PN WO200153312-A1.  
XX 26-JUL-2001.  
PD 26-DEC-2000; 2000WO-US34263.  
XX 21-JAN-2000; 2000US-0488725.  
PR 25-APR-2000; 2000US-0552317.  
PR 09-JUL-2000; 2000US-0598042.  
PR 19-JUL-2000; 2000US-0620312.  
PR 03-AUG-2000; 2000US-0653450.  
PR 14-SEP-2000; 2000US-0662191.  
PR 19-OCT-2000; 2000US-0693036.  
PR 29-NOV-2000; 2000US-0727344.  
XX  
PA (HYSE-) HYSEQ INC.  
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D,  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J,  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
XX WPI; 2001-442253/47.  
DR N-PSDB; AAI59123.  
XX Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -  
XX Example 4; SEQ ID NO 3112; 10078bp; English.  
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and  
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and

CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
SQ Sequence 374 AA;  
Query Match 100.0%; Score 1345; DB 22; Length 374;  
Best Local Similarity 100.0%; Pred. No. 3.7e-144;  
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MESKMGELPLDINIOEPRWDSTFLGRARHFTVTDPRLNLISGAQLEASRNIVQNYRAG 60  
Db 1 MESKMGELPLDINIOEPRWDSTFLGRARHFTVTDPRLNLISGAQLEASRNIVQNYRAG 60  
QY 61 VVTPGITEDQLMRAKYVYDSAFHPDTEGKVVLIGRMSAQVPMNMTTTGCMLEFYRKTPTV 120  
Db 61 VVTPGITEDQLMRAKYVYDSAFHPDTEGKVVLIGRMSAQVPMNMTTTGCMLEFYRKTPTV 120  
QY 121 VFWQWVQNSFNAIVNYSNRSGDPTITVRQLGTAYVSATTGAVATAGLKSCLKPLPLVG 180  
Db 121 VFWQWVQNSFNAIVNYSNRSGDPTITVRQLGTAYVSATTGAVATAGLKSCLKPLPLVG 180  
QY 181 RFVFPFAVAANANCINIPLMRQRELOVGIPVADEAGQRLGYSVTAAKOGIFQVVISRICMA 240  
Db 181 RFVFPFAVAANANCINIPLMRQRELOVGIPVADEAGQRLGYSVTAAKOGIFQVVISRICMA 240  
QY 241 IPAMAIPLIMDTLEKDFLK 261  
Db 241 IPAMAIPLIMDTLEKDFLK 261  
RESULT 6  
AAB60083  
ID AAB60083 standard; Protein; 374 AA.  
XX AAB60083;  
AC AAB60083;  
DT 28-MAR-2001 (first entry)  
XX Human transport protein TPPT-3.  
DE Human; transport protein; TPPT; transport disorder; metabolic disorder;  
KW neurological disorder; cardiovascular disorder; reproductive disorder;  
KW immune disorder; cancer.  
XX  
OS Homo sapiens.  
PN WO200078953-A2.  
XX 28-DEC-2000.  
PD 16-JUN-2000; 2000WO-US16668.  
PF 17-JUN-1999; 99US-0139923.  
PR 10-AUG-1999; 99US-0148177.  
PR 18-AUG-1999; 99US-0149357.  
PR 28-OCT-1999; 99US-0162287.  
XX  
PA (INCY-) INCYTE GENOMICS INC.  
XX Lal P, Yang J, Yue H, Hillman JL, Tang YT, Bandman O, Burford N,  
PI Baughn MR, Azimzai Y, Lu DAM, Au-Young J, Patterson C;  
XX WPI; 2001-041424/05.  
DR N-PSDB; AAF27703.  
XX

PT Isolated polypeptide with a human transport protein sequence is useful  
PT for the diagnosis, prevention and treatment of disorders associated  
PT with the immune, reproductive and cardiovascular systems -  
XX  
PS Claim 2; Page 108-109; 165pp; English.  
XX  
CC The present invention provides the protein and coding sequences for 43  
CC novel human transport proteins (designated TPPTS). These can be used in  
CC the diagnosis and treatment of transport, metabolic, neurological,  
CC reproductive, cardiovascular and immune disorders, and cell proliferative  
CC disorders such as cancer.  
XX  
SQ Sequence 374 AA;  
  
Query Match 100.0%; Score 1345; DB 22; Length 374;  
Best Local Similarity 100.0%; Pred. No. 3.7e-144;  
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MESKMGELPLDINIQEPRWDQSTFLGRARHFTVTDPNLLLSGAQLASRNIVQNYRAG 60  
DB 1 MESKMGELPLDINIQEPRWDQSTFLGRARHFTVTDPNLLLSGAQLASRNIVQNYRAG 60  
QY 61 VTPGITEDQLWRAKYVYDSAFHPDTEGKVVLLIGRMSAQVPMNMTITGCMLEFYRKPTV 120  
DB 61 VTPGITEDQLWRAKYVYDSAFHPDTEGKVVLLIGRMSAQVPMNMTITGCMLEFYRKPTV 120  
QY 121 VFQWVNQSFNAIVNYSNRSGDPTITVRQLGTAYVSATTGAVATALLGLKSLTKHLPLVVG 180  
DB 121 VFQWVNQSFNAIVNYSNRSGDPTITVRQLGTAYVSATTGAVATALLGLKSLTKHLPLVVG 180  
QY 181 RFVPFAVAANAANCINIPLMRQRELQVGIPVADAGQRLGYSVTAKQGIQVVISRICMA 240  
DB 181 RFVPFAVAANAANCINIPLMRQRELQVGIPVADAGQRLGYSVTAKQGIQVVISRICMA 240  
QY 241 IPMAIIPPLIMDTLEKKDFLK 261  
DB 241 IPMAIIPPLIMDTLEKKDFLK 261  
  
RESULT 7  
AAB41589  
ID AAB41589 standard; Protein; 251 AA.  
XX  
AC AAB41589;  
XX  
DT 08-FEB-2001 (first entry)  
XX  
DE Human ORFX ORF1353 polypeptide sequence SEQ ID NO:2706.  
XX  
KW Human; open reading frame; ORFX; detection; cyostatic; hepatotropic;  
KW vulnery; antiporiatic; antiparkinsonian; nootropic; neuroprotective;  
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; antinflammatory disease; coagulation;  
KW thrombosis; contraceptive.  
XX  
OS Homo sapiens.  
XX  
PN WO200058473-A2.  
XX  
PD 05-OCT-2000.  
XX  
PF 31-MAR-2000; 2000WO-US08621.  
XX  
PR 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.  
PR 05-APR-1999; 99US-0127728.  
PR 30-MAR-2000; 2000US-0540763.  
XX  
PA (CURA-) CURAGEN CORP.  
XX  
PI Shinkets RA, Leach M;  
XX  
DR WPI, 2000-602362/57.  
DR N-PSDB; AAC75798.  
XX  
PT Novel nucleic acids and peptides derived from open reading frame X,  
PT useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease -  
XX  
PS Claim 11; Page 1942-1943; 5507pp; English.  
XX  
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cyostatic; hepatotropic; vulnery;  
CC antiporiatic; antiparkinsonian; nootropic; neuroprotective;  
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
CC antinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
CC antithyroid; and antianaemic. The sequences are used for determining  
CC the presence of or predisposition to, or preventing or treating  
CC pathological conditions associated with an ORFX-associated disorder. The  
CC nucleic acids can be used to express ORFX proteins in gene therapy  
CC vectors. The proteins and nucleic acids may be used to treat cancers,  
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
CC nocturnal haemoglobinuria, antinflammatory disease; to enhance  
CC coagulation; to inhibit thrombosis; and as a contraceptive.  
XX  
SQ Sequence 251 AA;  
  
Query Match 94.3%; Score 1269; DB 21; Length 251;  
Best Local Similarity 98.0%; Pred. No. 9.2e-136;  
Matches 246; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 MESKMGELPLDINIQEPRWDQSTFLGRARHFTVTDPNLLLSGAQLASRNIVQNYRAG 60  
DB 1 MESKMGELPLDINIQEPRWDQSTFLGRARHFTVTDPNLLLSGAQLASRNIVQNYRAG 60  
QY 61 VTPGITEDQLWRAKYVYDSAFHPDTEGKVVLLIGRMSAQVPMNMTITGCMLEFYRKPTV 120  
DB 61 VTPGITEDQLWRAKYVYDSAFHPDTEGKVVLLIGRMSAQVPMNMTITGCMLEFYRKPTV 120  
QY 121 VFQWVNQSFNAIVNYSNRSGDPTITVRQLGTAYVSATTGAVATALLGLKSLTKHLPLVVG 180  
DB 121 VFQWVNQSFNAIVNYSNRSGDPTITVRQLGTAYVSATTGAVATALLGLKSLTKHLPLVVG 180  
QY 181 RFVPFAVAANAANCINIPLMRQRELQVGIPVADAGQRLGYSVTAKQGIQVVISRICMA 240  
DB 181 RFVPFAVAANAANCINIPLMRQRELQVGIPVADAGQRLGYSVTAKQGIQVVISRICMA 240  
QY 241 IPMAIIPPLIM 251  
DB 241 IPMAIIPPLIM 251  
  
RESULT 8  
ABG20175  
ID ABG20175 standard; Protein; 397 AA.  
XX  
AC ABG20175;  
XX  
DT 18-FEB-2002 (first entry)



XX Novel human diagnostic protein #20166.  
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR N-PSDB; AAS84362.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 20; SEQ ID NO 50534; 103bp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG0377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 397 AA;  
Query Match 93.3%; Score 1255; DB 22; Length 397;  
Best Local Similarity 89.6%; Pred. No. 7.4e-134;  
Matches 251; Conservative 0; Mismatches 7; Indels 22; Gaps 2;

QY 4 KMGEPLDINIOEPRWDSTFLGRARHFTVTDPRNLLISGAOLEASRNIVQNYRAGVVT 63  
DB 70 KMGEPLDINIOEPRWDSTFLGRARHFTVTDPRNLLISGAOLEASRNIVQNYRAGVVT 129  
QY 64 PGITEDQLMRAKYVYDSAFHPDTEGEKVVLLIGRMSAQVPMNMTITGCMLEFY----- 114  
DB 130 PGITEDQLMRAKYVYDSAFHPDTEGEKVVLLIGRMSAQVPMNMTITGCMLEFYRQSGKDEGH 189  
QY 115 -----RKTPTVFWQWVNSFNAINVNSNRS-GDTPITVRQLGTAVYSATTGA 161  
DB 190 CRGRSECLSLRKTPTVFWQWVNSFNAINVNSNRSRWHFPSLVRLGTAVYSATTGA 249  
QY 162 VATALGLSLTKHLPLVGRFVFPFAAANAANCINIPLMRQRELQVGIPVADEAGQRLGYS 221  
|||||

DB 250 VATALGLSLTKHLPLVGRFVFPFAAANAANCINIPLMRQRELQVGIPVADEAGQRLGYS 309  
QY 222 VTAAGQIGFQVVISRICMAIPAMAIPPLIMDTLEKDFLK 261  
DB 310 VTAAGQIGFQVVISRICMAIPAMAIPPLIMDTLEKDFLK 349  
RESULT 9  
AAM40413  
ID AAM40413 standard; Protein; 322 AA.  
XX  
AC AAM40413;  
XX  
DT 22-OCT-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 3558.  
XX  
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia.  
XX  
OS Homo sapiens.  
XX  
PN WO200153312-A1.  
XX  
PD 26-JUL-2001.  
XX  
PF 26-DEC-2000; 2000WO-US34263.  
XX  
PR 21-JAN-2000; 2000US-048725.  
PR 25-APR-2000; 2000US-0552317.  
PR 09-JUL-2000; 2000US-0598042.  
PR 19-JUL-2000; 2000US-0620312.  
PR 03-AUG-2000; 2000US-0653450.  
PR 14-SEP-2000; 2000US-0662191.  
PR 19-OCT-2000; 2000US-0693036.  
PR 29-NOV-2000; 2000US-0727344.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
XX  
DR WPI; 2001-442253/47.  
DR N-PSDB; AAI59569.  
XX  
PT Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -  
XX  
PS Example 8; SEQ ID NO 3558; 10078bp; English.  
XX  
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and  
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemia and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
SQ Sequence 322 AA;

	Query Match	81.0%;	Score 1089;	DB 22;	Length 322;	
	Best Local Similarity	78.5%;	Pred. No. 4.4e-115;			
	Matches 201;	Conservative	28;	Mismatches 27;	Indels 0;	Gaps 0;
QY	6 GELPLDINIOEPRWDSTFLGARHFFVTDPRLNLLSGAOLEASRNIVONYRAGVWTPG	65				
Dp	3 GELPNNNIKEPRWDSTFIGRANHFFVTDPRLNLLTNEQLSARKIVHDYRQGI VPPG	62				
QY	66 ITEDOLMRAKYVYDSAFHPDTGEKVLIGRMSAQVPNMNTITGCMLTFYRKPTPTVFVFWQ	125				
Dp	63 LTENELMRAKIYDSA FHPDTGEKNILLGRMSAQVPNMNTITGCMTFYRTTPAVLFWQW	122				
QY	126 VNQSFNALVNYSNRSGDPTI TVRQLGTAYVSATTGA VATALGLSKLTJHLPLVGRFVPF	185				
Dp	123 INQSFNAVNYTNRS GDAPLTVNELGTAYVSATTGA VATALGLNALTKHVSP LIGRFVPF	182				
QY	186 AAVAANAANCINI PLMRQRELQVGIPVAD EAGQR LGYSTAAKQGFQVVISRICMAI PAMA	245				
Dp	183 AAVAANAANCINI PLMRQRELKVGI PVT DENG NR LGESANNAKAQA ITOVVVSRI LMAAPGMA	242				
QY	246 IPPLIMDTLEKKDFLK 261					
Dp	243 IPPFIMNTLEKKAFLK 258					

RESULT	10
ID	ABB79572 standard; Protein; 322 AA.
AC	ABB79572;
DT	01-OCT-2002 (first entry)
DE	Alzheimer's disease associated protein ADPI-41.
KW	ADPI-4; human; neuropsychiatric disorder; neurological disorder; Alzheimer's disease; nootropic; neuroprotective; neuroleptic; antiparkinsonian; antidepressant; diagnosis; vaccine; human; sideroflexin 1; Sfxn1; gene therapy.
OS	Homo sapiens.
FH	Key
FT	Peptide
FT	/note= "tryptic peptide"
FT	56..70
FT	/note= "tryptic peptide"
FT	80..83
FT	/note= "motif conserved in mouse sideroflexin 1"
FT	95..120
FT	/note= "predicted transmembrane domain"
FT	119..136
FT	/note= "motif conserved in mouse sideroflexin 1"
FT	151..167
FT	/note= "predicted transmembrane domain"
FT	171..178
FT	/note= "tryptic peptide"
FT	175..196
FT	/note= "predicted transmembrane domain"
FT	203..215
FT	/note= "tryptic peptide"
FT	224..233
FT	/note= "tryptic peptide"
FT	229..250
FT	/note= "predicted transmembrane domain"
FT	265..286
FT	/note= "predicted transmembrane domain"
FT	306..314
FT	/note= "tryptic peptide"
PN	WO200246221-A2.

13-JUN-2002.  
10-DEC-2001; 2001WO-GB05459.  
08-DEC-2000; 2000US-254431P.  
(OXFO-) OXFORD GLYCOSCIENCES UK LTD.  
Herach HMAc, Parekh RB, Rohlf C, Terrett JA, Tyson KL;  
WPI; 2002-557575/59.  
N-PSDB; ABN84365.  
Isolated or recombinant polypeptide useful in the treatment of  
neuropsychiatric or neurological disorder comprises a specific amino  
acid sequence -  
Claim 3; Fig 2b; 51pp; English.  
The present sequence is the protein sequence of a human protein,  
named ADPI-41, which is differentially expressed (present in  
reduced amounts) in all areas of the brain tested except the  
amygdala in Alzheimer's disease (AD). ADPI-41 has an apparent  
mol.wt. of 32806 and pI of 9.84. It is useful as a marker and/or  
therapeutic target in AD. The amino acid sequence is 95% identical  
to that of sideroflexin 1 (Sxfl1), recently identified in the mouse  
model of sideroblastic anaemia. This is the first demonstration of  
the human Sxfl protein and the first indication that it may be  
associated with the occurrence of AD. ADPI-41 undergoes  
alternative splicing to result in an isoform of 261 amino acids  
(see ABB79573). The invention provides ADPI-41 polypeptides and  
polynucleotides, vectors, host cells and antibodies. A claimed  
method of screening for and/or diagnosis of a neuropsychiatric or  
neurological disorder in a subject and/or monitoring the  
effectiveness of therapy comprises detecting and/or quantifying  
the amount of ADPI-41 polypeptide and/or nucleic acid molecule in a  
sample from the subject. A claimed method for the prophylaxis  
and/or treatment of a neuropsychiatric disorder or neurological  
disorder comprises administering an ADPI-41 polypeptide or nucleic  
acid. A method of screening for agents that modulate ADPI-41  
activity or expression is also claimed. Disorders that may be  
treated or diagnosed include AD, vascular dementia, Lewy body  
dementia, schizophrenia, Parkinson's disease, multiple sclerosis,  
and depression.

	Query Match	81.0%;	Score 1089;	DB 23;	Length 322;	
	Best Local Similarity	78.5%;	Pred. No. 4.4e-115;			
	Matches	201;	Conservative	28;	Mismatches	27;
					Indels	0;
					Gaps	0;
QY	6 GELPLDINIOEPRWOSTFLGRARHFTVTDPRNLISGAQLASRNIVQNRYRAGVVTCG	65				
Dd	3 GELPNNINIKEPRWDOSTFIGRANHFTVTDPRNILLTNEQLSARKIVHDYRQGIYPBG	62				
QY	66 ITEDQLMRKAVVYDSAFHPDTGEKVLLIGRMSAQVPNMNTITGCMLTFYRKPTFVVFWQW	125				
Dd	63 ILENELMRAKIYDSAFHPDTGEKMILIGRMSAQVPNMNTITGCMTFYRTTPAVLFWQW	122				
QY	126 VNQSFNAINVNSNRSGDPIITVRQLGTAYVSATTGAVATALGLSLTKHLPLVGRFVBF	185				
Dd	123 INQSFNNAVNYTNRSGBPPLTVNELGTAYVSATTGAVATAALGNALTKHVSPILGRFVBF	182				
QY	186 AAVAANAANCINIPLMKRELOVGIPVADDEAGQRLGYSTAAKGQIFQVVISRICMAIPAMA	245				
Dd	183 AAVAANAANCINIPLMKRELEKVGIPVTDENGNRLGESANNAKAQAITQVVVSRIIMAAPGMA	242				
QY	246 IPPLIMDTLEKKDPLK	261				
Dd	243 IPPFIMNTLEKKAFLK	258				

## RESULT 11

ABG68039	ID	ABG68039	standard; Protein; 322 AA.
XX	AC	ABG68039;	
XX	DT	07-OCT-2002	(first entry)
XX	DE	Human Alzheimer's disease-associated protein isoform-41 (ADPI-41).	
XX	KM	Human; Alzheimer's disease; AD; brain tissue; ADF; ADPI;	
KM	KW	Alzheimer's disease-associated feature; neuroprotective;	
KW	KW	Alzheimer's disease-associated protein isoform; nootropic;	
XX	ADPI-41.		
XX	OS	Homo sapiens.	
XX	PN	WO200246767-A2.	
PD	XX	13-JUN-2002.	
XX	XX	29-NOV-2001; 2001WO-GB05289.	
PF	XX	08-DEC-2000; 2000US-254431P.	
PR	XX	(OXFO-) OXFORD GLYCOSCIENCES UK LTD.	
PA	XX	Herath HMAc, Parekh RB, Rohlf C;	
PI	XX	WPI; 2002-508575/54.	
XX	DR	N-PSDB; ABK96715.	
PT	XX	Screening, diagnosis or prognosis of Alzheimer's disease in subject,	
PT	PT	comprises detecting Alzheimer's disease-associated features or	
PT	PT	Alzheimer's disease-associated protein isoforms in brain tissue	
PT	PT	from the subject	
PS	PS	Example 2; Fig 2b; 427pp; English.	
XX	CC	The present invention relates to methods and compositions for the	
CC	CC	screening, diagnosis or prognosis of Alzheimer's disease (AD) in	
CC	CC	a subject. The method comprises analysing a sample of brain tissue	
CC	CC	from a subject by 2D electrophoresis to generate a 2D array of	
CC	CC	Alzheimer's disease-associated features (ADFs), whose relative	
CC	CC	abundance correlates with the presence, absence, stage or severity of	
CC	CC	AD and comparing the abundance of each feature with the abundance of	
CC	CC	that chosen feature in brain tissue from persons free from AD. The	
CC	CC	invention also describes Alzheimer's disease-associated protein	
CC	CC	isoforms (ADPIs) detectable in brain tissue. The methods and	
CC	CC	compositions of the invention are useful for the screening, diagnosis	
CC	CC	or prognosis of AD in a subject, for determining the stage or severity	
CC	CC	of AD in a subject, for identifying a subject at risk of developing AD,	
CC	CC	or for monitoring the effect of therapy administered to a subject	
CC	CC	having AD. Antibodies capable of binding to ADPIs are useful for	
CC	CC	treating or preventing AD, and for determining the efficacy of a given	
CC	CC	treatment regime. An agent that modulates the activity of ADPI is	
CC	CC	useful in the manufacture of a medicament for the treatment or	
CC	CC	prevention of AD in a subject. The present sequence represents human	
CC	ADPI-41.		
XX	SQ	Sequence 322 AA;	
XX	SQ		
QY	Query Match	81.0%; Score 1089; DB 23; Length 322;	
QY	Best Local Similarity	78.5%; Pred. No. 4.4e-115;	
QY	Matches 201; Conservative	28; Mismatches 27; Indels 0; Gaps 0;	
DB	6	GELPLDINIOEPRWDOSTFLGRARHFTVTDPRNLLSGAQLASRNIVQNYRAGVTPG	65
DB	3	GELPPNINIKERPWDOSTFGRANHFVTVDPRNILLTNEQLSARKIVHDYRGIVPG	62
QY	66	ITEDQLMRKVVYDSAFHPDTGKVVILIGMSAQVPMNMTTTGCMLTFYRKTPTVVFQW	125
QY	63	LTENELMRKVIYDSAFHPDTGKVVILIGMSAQVPMNMTTTGCMMTFYRTPAVLFWQW	122

```

QY 126 VNQSENAIVNYSNRSGDPTITVRQLGTAYVSATTGAVATATLGLKSLTKLPLVGRFVVF 183
CC :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 123 INQSFNAVVNNTNRSGDAPLTVELGTAYVSATTGAVATATLGLNALTKHVSPLIGRFVVF 182
QY 186 AAVAANCINIPLMRQRELQVGIVPADEAGQRLGYSVTAQKQIFQVVISRICMAIPAMA 245
CC ||||||||||||||||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 183 AAVAANCINIPLMRQRELKVGIFVTDENGNRLGESANAAKQAITQVVSRILMAAFGMA 242
QY 246 IPPLIMDTLEKDFLK 261
DB 243 IPPFIMNTLEKKAFLK 258

RESULT 12
ABB76448
ID ABB76448 standard; Protein; 322 AA.
XX
AC ABB76448;
XX
DT 02-SEP-2002 (first entry)
DE Human protein cluster I polypeptide.
XX
KW Protein cluster I; human; metabolic disorder; obesity; diabetes;
KM antidiabetic; diagnosis; therapy.
XX
OS Homo sapiens.
XX
PN WO200242324-A1.
XX
PD 30-MAY-2002.
XX
PF 22-NOV-2001; 2001WO-SE02581.
XX
PR 24-NOV-2000; 2000SE-0004325.
XX
PA (PHAA ) PHARMACIA AB.
XX
PI Altersand A;
XX
DR WPI; 2002-500277/53.
DR N-PSDB; ABN83757.
XX
PT Novel nucleic acid molecule encoding Protein Cluster I, useful in the
PT diagnosis of metabolic diseases, such as obesity and diabetes, and in
PT the identification of agents useful in the treatment of the diseases
XX
PS Claim 3; Page 32-33; 34pp; English.
XX
XX The present invention relates to the identification of a human gene
CC family (see ABN83754-57) encoding a group of polypeptides (see
CC ABB76445-48) referred to as Protein Cluster I. This family of
CC homologous proteins was identified by an 'all-versus-all' BLAST
CC procedure using all Caenorhabditis elegans proteins in the
CC Wormpep20 database release. Proteins of unknown function were
CC compared to the Drosophila melanogaster Flybase database, and
CC non-annotated protein clusters, conserved in both C. elegans and D.
CC melanogaster, were used in a BLAST procedure against the Celera
CC Human Genome Database, and Protein Cluster I proteins of unknown
CC function were selected for study. The human part of Protein
CC Cluster I comprises polypeptides encoded by 3 genes. The gene
CC encoding the present polypeptide is expressed primarily in the
CC immune system. The claimed Protein Cluster I nucleic acid
CC molecules and proteins are proposed to be useful for differential
CC identification of the tissue(s) or cell type(s) present in a
CC biological sample, for diagnosis of diseases and disorders,
CC including metabolic disorders and immune disorders, especially
CC obesity and diabetes, and for identifying agents useful in the
CC treatment of such diseases.
XX
SQ Sequence 322 AA;

```

	Best Local Similarity	78.5%;	Pred. No. 4.4e-115;	
	Matches	201;	Conservative	28; Mismatches 27; Indels 0; Gaps 0;
QY	6	GELPLDINIQEPRWDQSTFLGRARHFFVTVDPRNLLSGAQIEASRNIVQNRYAGVWTPG	65	
Db	3	GELPPNINIKEPRWDQSTFLGRANHFFVTVDPRNILLTNEQIESARKIVHDYRQGIWPPG	62	
QY	66	ITEDQLMRAKVVYDSA FHPDTGEKVLLIGRMSAQQVPNNMTITGCMLTFYRKPTPVWFQM	125	
Db	63	LTENELMRAKIYDSAFHPDTGEKMILLIGRMSAQQVPNNMTITGCMMTFYRTTPAVLFWQM	122	
QY	126	VNQSFNAINVYSNRSGDPITVRQLGTAYVSATGGAVATAIGLSLTKLPLPLVGRFPVF	185	
Db	123	INQSFNAVVNYNRSGDAPLTVNELGTAYVSATTGA VATALGINALTKHVSPLIGRFVPF	182	
QY	186	AAVA AANCINI PLMRQRELOVGI PVAD EAGORLGYSVTAAXOGIFQVVISRICMAI PAMA	245	
Db	183	AAVA AANCINI PLMRQRELKYGIPVTDENGNRLGESANAARQA ITQVVSRILMAAPGM A	242	
QY	246	IPP LIMDTLEKKDFLK	261	
Db	243	IPP FIMNTLEKKAF LK	258	

XX The invention relates to human transporters and ion channels (TRICH)  
CC and the polynucleotides encoding them. The composition comprising TRICH  
CC or agonist of TRICH is useful for treating a disease or condition  
CC associated with decreased expression of functional TRICH or condition  
CC associated with overexpression of TRICH respectively. The composition  
CC comprising Ab is useful for diagnosing a condition of disease associated  
CC with expression of TRICH in a subject, where the disorders include a  
CC transport disorder such as akinesia, cystic fibrosis, diabetes mellitus,  
CC Parkinson's disease, myasthenia gravis, cardiac disorders associated  
CC with transport e.g. angina, hypertension, myocarditis, neurological  
CC disorders associated with transport e.g. Alzheimer's disease, Wilson's  
CC disease, schizophrenia, cataracts, infertility, hyperglycaemia, Grave's  
CC disease, goitre, Addison's disease, Huntington's disease, dementia,  
CC multiple sclerosis, bacterial and viral meningitis. TRICH DNA is useful  
CC for generating a transcript image of a tissue or cell type, which  
CC represents the global pattern of gene expression by a particular tissue  
CC or cell type and for analysing the proteome of a tissue or cell type.  
CC TRICH DNA is used in gene therapy. The present amino acid sequence is  
CC human TRICH3 protein.  
XX

RESULT 13	
AAE13276	
ID	AAE13276 standard; Protein; 322 AA.
XX	
AC	AAE13276;
XX	
DT	12-FEB-2002 (first entry)
XX	
DE	Human transporters and ion channels (TRICH)-3.
XX	
KW	Human; transporter and ion channel; TRICH; akinesia; cystic fibrosis;
KW	diabetes mellitus; Parkinson's disease; myasthenia gravis; dementia;
KW	cardiac disorder; angina; hypertension; myocarditis; hyperglycaemia;
KW	neurological disorder; Alzheimer's disease; cataract; infertility;
KW	Wilson's disease; schizophrenia; Grave's disease; addison's disease;
KW	Huntington's disease; multiple sclerosis; meningitis; hypotensive;
KW	cardiant; nootropic; neuroprotective; neuroleptic; ophthalmological;
KW	antithyroid; anticonvulsant; goitre; antiinflammatory.
XX	
OS	Homo sapiens.
XX	
PN	WO200177174-A2.
XX	
PD	18-OCT-2001.
XX	
PF	06-APR-2001; 2001WO-US11206.
XX	
PR	06-APR-2000; 2000US-195595P.
PR	12-APR-2000; 2000US-196872P.
PR	20-APR-2000; 2000US-199020P.
PR	28-APR-2000; 2000US-200552P.
PR	05-MAY-2000; 2000US-202348P.
PR	11-MAY-2000; 2000US-203495P.
XX	
PA	(INCY-) INCYTE GENOMICS INC.
XX	
PI	Reddy R, Thornton M, Borowsky ML, Tang YT, Khan FA, Tribouley CM
PI	Gandhi AR, Yao MG, Sanjanwala MS, Baughn MR, Nguyen DB,
PI	Polickey JL, Yue H, Selhamer JJ, Walia NK, Lal P, Kearney L;
PI	Walsh RT, Lu DAM, Lu Y, Greene BD, Raumann BE, Patterson C;
XX	
DR	WPI; 2002-017448/02.
DR	N-PSDB; AAD21995.
XX	
PT	Polypeptides of human transporters and ion channels, useful for
PT	diagnosing, treating or preventing disorders of transport,
PT	neurological, muscle, immunological and cell proliferative disorders
XX	
PS	Claim 1; Page 120; 150pp; English.

Query Match	81.0%;	Score 1089;	DB 23;	Length 322;
Best Local Similarity	78.5%;	Pred. No. 4.4e-115;		
Matches 201;	Conservative 28;	Mismatches 27;	Indels 0;	Gaps
QY 6	GELPLDINIQEPRWDQSTFLGRARHFFVTTPRNLISGAQLEASRNIVQNYRAGVTPG	655		
DB 3	GELPKNINIKPRWDQSTFIGRANHFFVTTPRNLITNEQLSARKIVHDYRQGIYPG	622		
QY 66	ITEDQLMRKVVYDSAFHPDTEKEVLLIGRMSAQVPMNMNTITGCMLTFRYKTPFTVVFQW	122		
DB 63	LTENELMRKVIYDSAFHPDTEKEMILIGRMSAQVPMNMNTITGCMTFYRTTPAVLFQW	122		
QY 126	VNQSFAIVNYSNRSGDTPITVRQLGTAYVSATGAVATAGLKSLLTKHLPPLVGRFVPF	188		
DB 123	INQSFAVAVNTYRSGDAPLTVNELGTAYVSATGAVATAGLNALTKHVSPLIGRFVPF	188		
QY 186	AAVAAANCINIPLMRQRELQVGPVADAGQRLGYSVTAQKQIFQVVISRICMAIPMA	24		
DB 183	AAVAAANCINIPLMRQRELKVGIPVTDENGRIGESANAKQAITQVVSRIIMAAPMA	24		
QY 246	IPPLIMDTLEKKDFLK 261			
DB 243	IPPFIMNTLEKKAFK 258			
RESULT 14				
AAB41585				
ID	AAB41585	standard; Protein; 332 AA.		
XX	AC	AAB41585;		
XX	DT	08-FEB-2001 (first entry)		
DE	Human ORFX ORF1349	polypeptide sequence SEQ ID NO:2698.		
XX	Human; open reading frame; ORFX; detection; cytosstatic; hepatotropic;			
KW	vulnerable; antipsooriatic; antiparkinsonian; nootropic; neuroprotective;			
KW	anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;			
KW	immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;			
KW	hypotensive; dermatological; immunosuppressive; antinflammatory;			
KW	antiviral; antibacterial; antifungal; antirheumatic; antithyroid;			
KW	antianaemic; gene therapy; cancer; proliferative disorder; hypertension;			
KW	neurodegenerative disorder; osteoarthritis; graft vs host disease;			
KW	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;			
KW	cholesterol ester storage; systemic lupus erythematosus; infection;			
KW	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;			
KW	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;			
KW	bone damage; cartilage damage; antiinflammatory disease; coagulation;			
XX	thrombosis; contraceptive.			

PS Claim 1; Page 120; 150pp; English.

[illegible]



```
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shlmkets RA, Leach M;
XX
DR WPI; 2000-602362/57.
DR N-PSDB; AAC75794.
XX
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
XX
PS Claim 11; Page 1938; 5507pp; English.
XX
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antipariatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticoagulant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antihydral; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 332 AA;

Query Match 81.0%; Score 1089; DB 21; Length 332;
Best Local Similarity 78.5%; Pred. No. 4.7e-115;
Matches 201; Conservative 28; Mismatches 27; Indels 0; Gaps 0;

QY 6 GELPLDINIOEPRWDSTFLGRARHFTVTPRNLISGAQLASRNIVQNYRAGVTPG 65
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 13 GELPNNINIKPRWDSTFGRANHFTVTPRNLITNEQLSARKIVHROGIVPPG 72
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 66 ITEDQLWRAKYVDSAFHPDTGEKVLIGRMSAQVPMNMVITGCMLEFYRKPTVVFQW 125
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 73 LTENELWRAKYVDSAFHPDTGEKMLIGRMSAQVPMNMVITGCMLEFYRKPTVVFQW 132
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 126 VNQSFNAIVNYSNRSGPTITVRQLGTAYVSATGAVATAGLKLSTKALPLVGRFVPE 185
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 133 INQSFNAVNYNTRSGDAPLTVNELGTAYVSATGAVATAGLKLSTKALPLVGRFVPE 192
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 186 AAVAAANCINIPLMRQRELOVGIPIVADAGQRLGYSVTAKOGIFQVVISRICMAIPAMA 245
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 193 AAVAAANCINIPLMRQRELOVGIPIVADAGQRLGYSVTAKOGIFQVVISRICMAIPAMA 252
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 246 IPPIMDTLEKKDFLK 261
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 253 IPPFIMNTLEKKAFLEK 268
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
RESULT 15
AAM42198
ID AAM42198 standard; Protein; 351 AA.
XX
AC AAM42198;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 7129.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR N-PSDB; AAI61354.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Example 2; SEQ ID NO 7129; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 351 AA;

Query Match 81.0%; Score 1089; DB 22; Length 351;
Best Local Similarity 78.5%; Pred. No. 5.1e-115;
Matches 201; Conservative 28; Mismatches 27; Indels 0; Gaps 0;
```

QY 6 GELPLDINIQEPRWDQSTFLGRARHFTVTDPNRLLLSGAQLEASRNIVQNTYRAGVVTG 65  
Db 32 GELPNINIKEPRWDQSTFLGRANHFTVDPNRILITNEQLESARKIVHDYRQGI VPPG 91  
QY 66 ITEDQLMRAKYVYDSAFHPDTGKVVLLIGRMSAQVPMNMITGCMLTFFYRKTPTVVFQW 125  
Db 92 LTENELMRAKYIYDSAFHPDTGKMWLLIGRMSAQVPMNMITGCMMTFFYRTTPAVLFWQW 151  
QY 126 VNQSFNAIVNYSNRSGDPTITVRQLGTAYVSATTGAVATAGLKSLLTKHLPLVGRFVPF 185  
Db 152 INQSFNAVNYTNRSQDAPLTVNELGTAYVSATTGAVATAGLNLTKHVSPLIGRFVPF 211  
QY 186 AAVAAANCINIPLMRQRELOYGIPVADAGORLGYSVTAAKQGIPOVVISRICMAIPAMA 245  
Db 212 AAVAAANCINIPLMRQRELKVGIPVTDENGRLGESANAAKQAITOVVVSRIIMAAPGMA 271  
QY 246 IPPLIMDTLEKDFLK 261  
Db 272 IPPFIMNTLEKKAFLK 287

Search completed: February 3, 2004, 09:12:44  
Job time : 42 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 3, 2004, 09:11:53 ; Search time 21 Seconds

(without alignments)  
525.864 Million cell updates/sec

Title: US-09-990-415A-2

Perfect score: 1345  
Sequence: 1 MESKMGELPLDINIQEPRWD.....PAMAIPLIMDTLEKKDLK 261

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	96	7.1	602	2 US-08-419-652-6	Sequence 6, Appli
2	96	7.1	783	6 5422248-2	Patent No. 5422248
3	96	7.1	836	1 US-07-923-976-4	Sequence 4, Appli
4	89.5	6.7	863	1 US-07-923-976-8	Sequence 8, Appli
5	89	6.6	655	4 US-09-134-001C-5531	Sequence 5531, Ap
6	88	6.5	394	4 US-09-252-991A-18502	Sequence 18502, A
7	82	6.1	146	2 US-08-627-173-20	Sequence 20, Appli
8	82	6.1	146	2 US-08-535-882A-20	Sequence 20, Appli
9	82	6.1	146	3 US-08-316-424A-6	Sequence 6, Appli
10	82	6.1	146	3 US-09-005-546-20	Sequence 20, Appli
11	82	6.1	146	4 US-08-477-669-6	Sequence 6, Appli
12	82	6.1	771	1 US-07-923-976-6	Sequence 6, Appli
13	81	6.0	514	4 US-09-252-991A-25845	Sequence 25845, A
14	81	6.0	676	3 US-08-947-965-71	Sequence 71, Appli
15	80	5.9	621	4 US-09-252-991A-19749	Sequence 19749, A
16	80	5.9	815	4 US-09-914-259-18	Sequence 18, Appli
17	79.5	5.9	529	5 PCT-US92-00282-7	Sequence 7, Appli
18	79.5	5.9	693	1 US-08-463-620-11	Sequence 11, Appli
19	79.5	5.9	693	2 US-08-224-917-11	Sequence 11, Appli
20	79.5	5.9	693	2 US-08-914-853-11	Sequence 11, Appli
21	79.5	5.9	693	5 PCT-US95-03934A-11	Sequence 11, Appli
22	77	5.7	435	4 US-09-252-991A-17750	Sequence 17750, A
23	76	5.7	1209	4 US-09-252-991A-25844	Sequence 25844, A
24	75.5	5.6	531	5 PCT-US92-00282-5	Sequence 5, Appli
25	75	5.6	552	4 US-09-773-426A-9	Sequence 9, Appli
26	75	5.6	627	4 US-09-252-991A-23072	Sequence 23072, A
27	75	5.6	705	4 US-09-252-991A-27442	Sequence 27442, A

28	74	5.5	430	4 US-08-311-731A-153	Sequence 153, App
29	74	5.5	654	4 US-09-134-001C-3261	Sequence 3261, Ap
30	73.5	5.5	454	1 US-07-915-720D-17	Sequence 17, Appl
31	73.5	5.5	454	3 US-09-025-543-17	Sequence 17, Appl
32	73.5	5.5	456	1 US-07-915-720D-19	Sequence 19, Appl
33	73.5	5.5	456	3 US-08-879-565-11	Sequence 11, Appl
34	73.5	5.5	456	3 US-09-025-543-19	Sequence 19, Appl
35	73.5	5.5	456	3 US-08-431-517F-16	Sequence 16, Appl
36	73.5	5.5	487	1 US-08-030-644-2	Sequence 2, Appli
37	73.5	5.5	487	1 US-08-013-801-2	Sequence 2, Appli
38	73.5	5.5	487	1 US-08-072-063-2	Sequence 2, Appli
39	73.5	5.5	487	1 US-08-212-132-2	Sequence 2, Appli
40	73.5	5.5	487	1 US-08-414-924-2	Sequence 2, Appli
41	73.5	5.5	487	1 US-08-311-611A-69	Sequence 69, Appl
42	73.5	5.5	487	1 US-08-311-611A-146	Sequence 146, App
43	73.5	5.5	487	1 US-08-173-968-2	Sequence 2, Appli
44	73.5	5.5	487	1 US-08-232-527-2	Sequence 2, Appli
45	73.5	5.5	487	1 US-08-372-783-69	Sequence 69, Appl

## ALIGNMENTS

RESULT 1  
US-08-419-652-6  
; Sequence 6, Application US/08419652  
; Patent No. 5831007  
; GENERAL INFORMATION:  
; APPLICANT: Chua, Anne O  
; APPLICANT: Gubler, Ulrich A  
; TITLE OF INVENTION: INTERLEUKIN-12 RECEPTOR  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hoffmann-La Roche Inc.  
; STREET: 340 Kingsland Street  
; CITY: Nutley  
; STATE: New Jersey  
; COUNTRY: United States of America  
; ZIP: 07110-1199  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: PC compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/419,652  
; FILING DATE: 11-APR-1995  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/248,532  
; FILING DATE: 31-MAY-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/094,713  
; FILING DATE: 19-JUL-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kaas, Alan P  
; REGISTRATION NUMBER: 32142  
; REFERENCE/DOCKET NUMBER: CD 9174  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (201) 235-4205  
; TELEFAX: (201) 235-3500  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 602 amino acids  
; TYPE: amino acid  
; TOPOLOGY: 1linear  
; MOLECULE TYPE: protein  
; FEATURE:  
; NAME/KEY: Region  
; LOCATION: 1..602  
; OTHER INFORMATION: /note= "Represents residues 98 to  
; OTHER INFORMATION: 731 of human granulocyte colony-stimulating  
; OTHER INFORMATION: factor-receptor."





```

1  TITLE OF INVENTION:  Colony-Stimulating Factor Receptor
2  NUMBER OF SEQUENCES:  8
3  CORRESPONDENCE ADDRESS:
4  ADDRESSEE:  Jones, Tullar & Cooper, P.C.
5  STREET:  P.O. Box 2266 Eads Station
6  CITY:  Arlington
7  STATE:  Virginia
8  ZIP:  22202
9  COMPUTER READABLE FORM:
10 MEDIUM TYPE:  Floppy disk
11 COMPUTER:  IBM PC compatible
12 OPERATING SYSTEM:  PC-DOS/MS-DOS
13 SOFTWARE:  Patentin Release #1.0, Version #1.25
14 CURRENT APPLICATION DATA:
15 APPLICATION NUMBER:  US/07/923,976
16 FILING DATE:  19920922
17 CLASSIFICATION:  435
18 PRIOR APPLICATION DATA:
19 APPLICATION NUMBER:  JP 74539/1990
20 FILING DATE:  23-MAR-1990
21 PRIOR APPLICATION DATA:
22 APPLICATION NUMBER:  JP 176629/1990
23 FILING DATE:  03-JUL-1990
24 PRIOR APPLICATION DATA:
25 APPLICATION NUMBER:  PCT/JP91/00375
26 FILING DATE:  22-MAR-1991
27 ATTORNEY/AGENT INFORMATION:
28 NAME:  Helliwege, James W.
29 REGISTRATION NUMBER:  28,808
30 REFERENCE/DOCKET NUMBER:  514853
31 TELECOMMUNICATION INFORMATION:
32 TELEPHONE:  703-415-1500
33 TELEFAX:  703-415-1508
34 INFORMATION FOR SEQ ID NO:  8:
35 SEQUENCE CHARACTERISTICS:
36 LENGTH:  863 amino acids
37 TYPE:  AMINO ACID
38 TOPOLOGY:  linear
39 MOLECULE TYPE:  protein
40 US-07-923-976-8

```

```

Query Match      6.7%; Score 89.5; DB 1; Length 863;
Best Local Similarity 22.4%; Pred. No.0.29;
Matches 55; Conservative 35; Mismatches 90; Indels 65; Gaps 13;

QY          61 VVTPGITEDQLMRAK--YYVDSAFHPDTEGEKVVL--IGRMSAQ---VPMNMITITGCMLTF 113
           :|||::||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB          503 IVTP-LYQDTMGPSQHVVAYVSQEMAPSHAPBLHLKHIGKTWAQLEWVPBPPELGKSPLTH 561
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||

QY          114 YRKPTPVFW-QWVNQSFNALVNYSNRS----GDTPTT--VRQIGTAIVSATTGAVATA 165
           |::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB          562 Y-----TIFWTNAQNQSFSAILNASSRGFVLHGLEPASLYIHLMASQAAGATNSTVLTL 616
           ::|||::|||::|||::|||::|||::|||::|||::|||::|||

QY          166 LGLKSLTKHLPPLVGFRFPFAVAANCI-----NIPLMRQRELQVG 207
           :|||::|||::|||::|||::|||::|||::|||::|||::|||
DB          617 MTLTPGESELHIILGL--FGILLLLLTCLGTAWLCCSPNRKNPDLMPSPDPDAHSLSLGSW 673
           |||::|||::|||::|||::|||::|||::|||::|||::|||

QY          208 IPVADE-----AGORLGYSVTAAK-----OGIFOVVISRICMAIPAMAIPLLIMDT 253
           :|||::|||::|||::|||::|||::|||::|||::|||::|||
DB          674 VPTIMEELPGPRQGWLGGTSEMSRALTPHPCVDQAFQ-----LPGLGTPPIITKLTT 724
           :|||::|||::|||::|||::|||::|||::|||::|||::|||

QY          254 LEKD 258
           :::|
DB          725 VLED 729

RESULT 5
US-09-134-001C-5531
; Sequence 5531, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
```

```

; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5531
; LENGTH: 655
; TYPE: PR1
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5531

```

Query Match	6.6%	Score 89	DB 4	Length 655
Best Local Similarity	20.9%	Pred. No. 0.21		
Matches	54	Conservative	32	Mismatches 78
				Indels 94
				Gaps 10
QY	53	IVQNYRAGVTPGITEBQLWRKAYVYDSAFHPDTEKVVILGHSQAQVPMNMTITGCMLT	112	
Db	407	IIAGFLAGYLTTQGI-----KYITRKLPAQIEGLKPTLIYPL-----LSVSITGLLMV	453	
QY	113	FYRKPTPVFMQWVNGSFNAIVNYSNRSGDPTITVRQL-----GTAYVSA	157	
Db	454	YVENPPAA---WLN---HLLNGLNSLSGSNIMLGLVIGAMMAIDMGGFENKAYVFA	506	
QY	158	TTG-----AVATA-----LGLSLTKHL	175	
Db	507	TAALTEGNAPITAMIGMIPPLAIIATAMLIFRKFTKEQRGSIVPNVYMGLSFITEGA	566	
QY	176	PPLVGRFVPPAAVAAANCINIPLMRQRELOVGIPVADGAGRLGYSVTAAKQIGIFOVV--	233	
Db	567	-----IPFAAAD-----PLRVIIPSMVSGVAGAIALGLSSIKAPHGIGFVIITGT	612	
QY	234	-ISRICMAIPAMAIPPLI	250	
Db	613	DFNHILQTLIALVVGTLV	630	

```

RESULT 6
US-09-252-991A-18502
; Sequence 18502, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18502
; LENGTH: 394
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18502

```

```
Query Match          6.5%; Score 88; DB 4; Length 394;
Best Local Similarity 18.7%; Pred. No. 0.12;
Matches 46; Conservative 40; Mismatches 64; Indels 96; Gaps 11;
```

OY      35 TDPRNLLSGAQLASRNIVQNYRAGVTPGTEDQLMRKVVYDSA----- 81  
          :::         : | | | | | | | | |  
DB      139 SEPLVQIVTDGSIPTNTANYVENTARYGV-----QFWRAG--LDVAAPQAQVMLEPRYW 189

OY      82 FHPDTEKEKVLLIGMSAQVPMMNTITGCMLTFYRKPTPVFWQMWNQSFNALVNYSNRSG 141  
          | : | : | |         : | | | | | | | | | : | : | :  
DB      190 FNPELESRRALI---PGAIAI VMTII GTM L-----TALVVARBERGMTAEVL S----- 235

QY 142 DTPTVRQLGTAVVSATTGAVATATLGLSLTKHLPLVGRFVPFAAV-----AAANCI 194  
DB 236 -TPASVAEL-----LIGKLPYFVLGMLSTLGAALAV 267  
QY 195 NIFLMRQRELOVGI PVADAGQRLGYSVFAAKOGIFQVVISRICMAIPAMAIPLIMDTL 254  
DB 268 FV-----FGVPM-----RGSLLSLLLSAVFWPALG-QGLLISL 302  
QY 255 EKXDFL 260  
DB 303 ARNOFL 308

## RESULT 7

US-08-627-173-20  
; Sequence 20, Application US/08627173  
; Patent No. 5861483  
; GENERAL INFORMATION:  
; APPLICANT: TSYRLOVA, IRENA  
; APPLICANT: WOLPE, STEPHEN D.  
; TITLE OF INVENTION: INHIBITOR OF STEM CELL PROLIFERATION AND  
; TITLE OF INVENTION: USES THEREOF  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NIXON & VANDERHVE P.C.  
; STREET: 1100 NORTH GLEBE ROAD  
; CITY: ARLINGTON  
; STATE: VIRGINIA  
; COUNTRY: U.S.A.  
; ZIP: 22201-4714  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/627,173  
; FILING DATE: 03-APR-1996  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/535,882  
; FILING DATE: 28-SEP-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BYRNE, THOMAS E.  
; REGISTRATION NUMBER: 32,205  
; REFERENCE/DOCKET NUMBER: 1331-177  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 816-4000  
; TELEFAX: (703) 816-4100  
; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 146 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-627-173-20

Query Match 6.1%; Score 82; DB 2; Length 146;  
Best Local Similarity 25.8%; Pred. No. 0.13;  
Matches 32; Conservative 19; Mismatches 57; Indels 16; Gaps 4;

QY 120 VVFWQWVNSFNATVNSNRSGDPTITVRQLGTAVVSATTGAVATATL--GLKSLTKHLPP 177  
DB 32 LVVYPMTQRYFDSF-----GDLSSASAIMGNKVKAHGKVITAFNDGL----NHLDS 80  
QY 178 LVGRFVPPAAVAANCINIPLMRQRELOVGI PVADAGQRLGYSVFAAKOGIFQVVISRI 237  
DB 81 LKG---TFASLSLHCDKLHVDPENFRLLGNMIVIVLGHHLGKDFTPAAQAFAQKVAVAGV 137  
QY 238 CMAI 241

DB 138 ATAL 141

## RESULT 8

US-08-535-882A-20  
; Sequence 20, Application US/08535882A  
; Patent No. 5939391  
; GENERAL INFORMATION:  
; APPLICANT: TSYRLOVA, IRENA  
; APPLICANT: WOLPE, STEPHEN D.  
; TITLE OF INVENTION: INHIBITOR OF STEM CELL PROLIFERATION AND  
; TITLE OF INVENTION: USES THEREOF  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NIXON & VANDERHVE P.C.  
; STREET: 1100 NORTH GLEBE ROAD  
; CITY: ARLINGTON  
; STATE: VIRGINIA  
; COUNTRY: U.S.A.  
; ZIP: 22201-4714  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/535,882A  
; FILING DATE: 28-SEP-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BYRNE, THOMAS E.  
; REGISTRATION NUMBER: 32,205  
; REFERENCE/DOCKET NUMBER: 1331-177  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 816-4000  
; TELEFAX: (703) 816-4100  
; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 146 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-535-882A-20

Query Match 6.1%; Score 82; DB 2; Length 146;  
Best Local Similarity 25.8%; Pred. No. 0.13;  
Matches 32; Conservative 19; Mismatches 57; Indels 16; Gaps 4;

QY 120 VVFWQWVNSFNATVNSNRSGDPTITVRQLGTAVVSATTGAVATATL--GLKSLTKHLPP 177  
DB 32 LVVYPMTQRYFDSF-----GDLSSASAIMGNKVKAHGKVITAFNDGL----NHLDS 80  
QY 178 LVGRFVPPAAVAANCINIPLMRQRELOVGI PVADAGQRLGYSVFAAKOGIFQVVISRI 237  
DB 81 LKG---TFASLSLHCDKLHVDPENFRLLGNMIVIVLGHHLGKDFTPAAQAFAQKVAVAGV 137  
QY 238 CMAI 241  
DB 138 ATAL 141

## RESULT 9

US-08-316-424A-6  
; Sequence 6, Application US/08316424A  
; Patent No. 6022848  
; GENERAL INFORMATION:  
; APPLICANT: KOZLOV, VLADIMIR  
; APPLICANT: TSYRLOVA, IRENA  
; TITLE OF INVENTION: INHIBITOR OF STEM CELL PROLIFERATION AND  
; TITLE OF INVENTION: USES THEREOF  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:



INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 146 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
US-08-477-669-6

Query Match 6.1%; Score 82; DB 4; Length 146;  
Best Local Similarity 25.8%; Pred. No. 0.13;  
Matches 32; Conservative 19; Mismatches 57; Indels 16; Gaps 4;

QY 120 VVFWQWQNSFNAIVNYSNRSGDPTITVRLGTAVYSATTGAVATAL--GLKSLTKLPP 177  
DB 32 LVVYPMWQRYFDSF-----GDLSSASAIMGNKAKHKKVITPAFNDGL---NHLDS 80

QY 178 LVGRFVPPAAVAANNCINIPLMRQRELVGIPVADEAGRLGYSVTAAKQGIPOVISRI 237  
DB 81 LKG---TFASLSEHLCHDKLHVDPENFRLLGNMIVIVLGHHLGKDFTPAAQAFAQKVAGV 137

QY 238 CMAI 241  
DB 138 ATAL 141

RESULT 12  
US-07-923-976-6  
Sequence 6, Application US/07923976  
Patent No. 5574136  
GENERAL INFORMATION:  
APPLICANT: Nagata, Shigekazu  
TITLE OF INVENTION: DNA Encoding Granulocyte  
TITLE OF INVENTION: Colony-Stimulating Factor Receptor  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Jones, Tullar & Cooper, P.C.  
STREET: P.O. Box 2266 Bads Station  
CITY: Arlington  
STATE: Virginia  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/923,976  
FILING DATE: 19920922  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 74539/1990  
FILING DATE: 23-MAR-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 176629/1990  
FILING DATE: 03-JUL-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP91/00375  
FILING DATE: 22-MAR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Hellwege, James W.  
REGISTRATION NUMBER: 28,808  
REFERENCE/DOCKET NUMBER: 514853  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-415-1500  
TELEFAX: 703-415-1508  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 771 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear

MOLECULE TYPE: protein  
US-07-923-976-6

Query Match 6.1%; Score 82; DB 1; Length 771;  
Best Local Similarity 29.1%; Pred. No. 1.9;  
Matches 41; Conservative 21; Mismatches 57; Indels 22; Gaps 9;

QY 61 VVTPGITEDQLMRK--YVYDSAFHPDTEKVL--IGRMSAQ---VPMNMTITGMLTF 113  
DB 503 LVTP-LYQDTMGPSQHVYVYSGEMAFSHAPBLHKHIGKWAQLEWVPEPPELGKPLTH 561

QY 114 YRKPTVFW-QWVNSFNAIVNYSNRSGDPTITVRLGTAVYSATTGAVATALGL 168  
DB 562 Y-----TIFWTAQNSFSALINASSRGFVLHGLEPASLYHIHLMAASQAGATNSTVLT 616

QY 169 KSLTKLPLLVGRFVPPAAVA 189  
DB 617 MTLT---PAPGR-IPSGQVS 633

RESULT 13  
US-09-252-991A-25845  
Sequence 25845, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
PRIOR FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 25845  
LENGTH: 514  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-25845

Query Match 6.0%; Score 81; DB 4; Length 514;  
Best Local Similarity 23.5%; Pred. No. 1.3;  
Matches 35; Conservative 24; Mismatches 52; Indels 38; Gaps 6;

QY 118 PTVFWQWVNSFNA-----IVNYSNRSGDT--PITVRQLGTAVYSATTGAVATAGLSL 171  
DB 376 PALAMFAVPLVFGAMNALLARYAAGDTRPLRCELLGSLNAAALLAVLPLIFGLPG- 434

QY 172 TKHLPPLVGRFVPPAAVAANNCINIPLMRQRELVGIPVADEAGRLGYSVTAAKQGIPO 231  
DB 435 -----IPLAALGVLTCTALLLMRQALLGALPWA-----RL-WLSALAMALAA 477

QY 232 VVISRI-----CMAIPAMAI 246  
DB 478 GLUFRIDGIWLQLGLGTVAGCIALLGMAI 506

RESULT 14  
US-08-947-965-71  
Sequence 71, Application US/08947965A  
Patent No. 6004790  
GENERAL INFORMATION:  
APPLICANT: Dijkhuizen, Lubbert  
APPLICANT: Dijkstra, Bauke  
APPLICANT: Andersen, Carsten  
APPLICANT: Osten, Claus von der  
TITLE OF INVENTION: Cyclomalto-dextrin Glucanotransferase  
TITLE OF INVENTION: Variants  
FILE REFERENCE: 4285.204-US  
CURRENT APPLICATION NUMBER: US/08/947,965A  
CURRENT FILING DATE: 1997-10-09



EARLIER APPLICATION NUMBER: 0477/95  
EARLIER FILING DATE: 1995-04-21  
EARLIER APPLICATION NUMBER: 1173/95  
EARLIER FILING DATE: 1995-10-17  
EARLIER APPLICATION NUMBER: 1281/95  
EARLIER FILING DATE: 1995-11-16  
EARLIER APPLICATION NUMBER: PCT/DK96/00179  
EARLIER FILING DATE: 1996-04-22  
NUMBER OF SEQ ID NOS: 78  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 71  
LENGTH: 676  
TYPE: PRT  
ORGANISM: Bacillus sp.  
US-08-947-965-71

Query Match 6.0%; Score 81; DB 3; Length 676;  
Best Local Similarity 20.5%; Pred. No. 2;  
Matches 43; Conservative 33; Mismatches 82; Indels 52; Gaps 9;

QY 67 TEDQLWRKYYVDSAFHPDTGEKVLIGMSAQVPMNTIT----- 107  
DB 405 TERWLNEIDYIYERTF---GNSIVLTAVNSN--SNQITINLNTSLPQGNVTDLOQL 458  
QY 108 -----GCMLEFYRKPTPTVFWQWVNSFNAIVN---YSNRSGDTPITVR----- 148  
DB 459 DGNITTVNANGAVNSFQLRANSVAVWQVNSPSTSPILQGVPMGKAGNT-ITVSGEGFG 517  
QY 149 -QLGTAYVSATGAVATAGLKLTKLPLPLVGRFVPFAVAANAANCINIPLMRQRELVG 207  
DB 518 DERGSVLFDSISSEIISWSNTKISVK-VPNVAGGYDLSVTAAN-IKSPYKEFEVLSG 575  
QY 208 IPVADEAGQRLGYSVTAQKQIFQVVISRI 237  
DB 576 ---NQSVRFQVNNATTSPGTNLYIVGNV 601

RESULT 15  
US-09-252-991A-19749  
Sequence 19749, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 19749  
LENGTH: 621  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-19749

Query Match 5.9%; Score 80; DB 4; Length 621;  
Best Local Similarity 21.9%; Pred. No. 2.3;  
Matches 39; Conservative 26; Mismatches 53; Indels 60; Gaps 9;

QY 88 EKVLLIGMSAQVPMNTITGCMLEFYRKPTPTVFWQWVNSFNAIVNSNRSGDTPITV 147  
DB 287 DAIVLGRVLASAPSVMAVGFASTYSSRTAE---QWQDQ-----TPDSV 328  
QY 148 RQLGTAYVSATGAVATAGLKLTKLPLPLVGRFVPFAVA-AANCINIPLMRQRELVG 205  
DB 329 RY-----ALGMDAKLGLPSPVN---LNAAKASGTVLPMRLTNBARG 369  
QY 206 -----VGIPVADEAGQRLGYSVTAQKQIFQVVISRICMAIPMAIIPPLIM 251

DB 370 NTTTLSTVSTDGVSVPKAVPV-RMAAYNATT---GLYEVTVP-----TTAEAPPLIL 418

Search completed: February 3, 2004, 09:15:15  
Job time : 23 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 3, 2004, 09:11:53 ; Search time 20 Seconds  
(without alignments)  
1255.001 Million cell updates/sec

Title: US-09-990-415A-2  
Perfect score: 1345  
Sequence: 1 MESKMGELPLDINIQEPRWD.....PAMAIPLIMDTLEKKDPLK 261

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 76:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1064	79.1	357	2 I55210	tricarboxylate car
2	742.5	55.2	324	2 T24465	hypothetical prote
3	693	51.5	326	2 T19996	hypothetical prote
4	639	47.5	329	2 T18612	hypothetical prote
5	604	44.9	326	2 T21924	hypothetical prote
6	593	44.1	329	2 T27337	hypothetical prote
7	483.5	35.9	285	2 T15498	hypothetical prote
8	464.5	34.5	327	2 S67168	probable membrane
9	432.5	32.2	616	2 T19873	hypothetical prote
10	411.5	30.6	325	2 T37847	probable transport
11	278	20.7	326	2 T02841	mitochondrial tric
12	101	7.5	791	2 A99514	hypothetical prote
13	96	7.1	783	2 JH0329	granulocyte colony
14	94.5	7.0	412	2 T40295	fructosyl amine -
15	89.5	6.7	344	2 I56551	neurotrophin - rat
16	89.5	6.7	863	2 C38252	granulocyte colony
17	87.5	6.5	234	2 B69437	hypothetical prote
18	87.5	6.5	569	2 G95418	probable Kdpa pote
19	87	6.5	343	2 AC2129	iron(III) diclrat
20	86	6.4	436	2 AB0645	ABC transporter in
21	85	6.3	458	2 S75375	hypothetical prote
22	85	6.3	914	1 S07047	iodide peroxidase
23	85	6.3	926	1 OPGIT	iodide peroxidase
24	84	6.2	291	2 T36190	probable transfera
25	84	6.2	398	2 F86640	multidrug resistan
26	83.5	6.2	469	2 B83368	hypothetical prote
27	83.5	6.2	503	2 S52377	sucrose transport
28	83.5	6.2	2591	2 T30288	pristinamycin I by
29	82.5	6.1	274	2 S50190	phosphotransferase

30	82.5	6.1	519	2 S31136	nitrogen fixation
31	82.5	6.1	625	2 G96976	probable periplasm
32	82	6.1	147	1 HBMS	hemoglobin beta ma
33	82	6.1	382	2 AF0168	probable membrane
34	82	6.1	771	2 B38252	granulocyte colony
35	82	6.1	837	2 S30971	gene 26 protein -
36	82	6.1	1077	2 B82566	acriflavin resista
37	81.5	6.1	531	2 A55788	glucuronosyltransf
38	81.5	6.1	939	2 T18974	hypothetical prote
39	81.5	6.1	1626	2 A75613	hypothetical prote
40	81	6.0	324	2 F70609	probable exported
41	81	6.0	328	2 AE0491	hypothetical prote
42	81	6.0	703	1 ALBSX1	cyclomaltoextrin
43	81	6.0	914	1 JN0550	iodide peroxidase
44	81	6.0	2505	1 XYRTFA	enoyl-lacyl-carrie
45	80.5	6.0	356	2 AG0370	probable ABC-trans

ALIGNMENTS

RESULT 1  
I55210  
tricarboxylate carrier - rat (fragment)  
C/Species: Rattus sp. (rat)  
C/Date: 26-Jul-1996 #sequence\_revision 26-Jul-1996 #text\_change 26-May-2000  
C/Accession: I55210  
R/Azzi, A.; Glerum, M.; Koller, R.; Mertens, W.; Spycher, S.  
J. Bioenerg. Biomembr. 25, 515-524, 1993  
A/Title: The mitochondrial tricarboxylate carrier.  
A/Reference number: I55210; MUID:94179133; PMID:8132491  
A/Accession: I55210  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-357 <RES>  
A/Cross-references: GB:S70011; NID:g545997; PIDN:AAB30258.1; PID:g545998  
C/Superfamily: Saccharomyces probable membrane protein YOR271c

Query Match 79.1%; Score 1064; DB 2; Length 357;  
Best Local Similarity 77.3%; Pred. No. 1.2e-85;  
Matches 198; Conservative 27; Mismatches 31; Indels 0; Gaps 0;

QY	6	GELPLDINIQEPRWDSTFLGRARHFTVTDPNLLLSGAQLASRNIVQNYRAGVTPG	65
DB	38	GEVPPNINIKPEPRWDQSTFIGRAHFTVTDPKNILITNEQLENAKVVDYRQGIYPAG	97
QY	66	ITEDQLRAKYVYDSAFHPDTGEKVVLIGRMSAQVPMNTITGCMITFRKPTVFWQW	125
DB	98	LTENELWRAKYAYDSAFHPDTGEKVVLIGRMSAQVPMNTITGCMITFRKPTVFWQW	157
QY	126	VNQSFNAINVSNRSGDPTITVROLGTAIVSATGAVATAGLSLTKHLPLLVGRFVPF	185
DB	158	INQSFNAINVSNRSGDPTITVROLGTAIVSATGAVATAGLSLTKHLPLLVGRFVPF	217
QY	186	AAVAAANCINIPLMRQRELQVGPVADAGQRLGYSVTAAKQGIPOVVISRICMAIPAMA	245
DB	218	AAVAAANCINIPLMRQRELQVGPVADAGQRLGYSVTAAKQGIPOVVISRICMAIPAMA	277
QY	246	IPPLIMDTLEKKDPLK 261	
DB	278	IPPLIMDTLEKKDPLK 293	

RESULT 2  
T24465  
hypothetical protein T04F8.1 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
C/Accession: T24465  
Riennard, N.  
submitted to the EMBL Data Library, November 1995  
A/Reference number: Z19895  
A/Accession: T24465



A;Introns: 26/1; 99/3; 171/3; 200/2; 243/1; 298/3  
C;Superfamily: Saccharomyces probable membrane protein YOR271c

Query Match 44.9%; Score 604; DB 2; Length 326;

Best Local Similarity 47.3%; Pred. No. 2.3e-45;

Matches 121; Conservative 47; Mismatches 86; Indels 2; Gaps 1;

```
QY      1 MESKMGEPLDINIDEPKWDQSTFLGRARHFFVTDPNNLLSGAQLASRNIVQNYRAG 60
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      1 MSELVSNLKTTRPDISKQWQRTYYGRVRFHFTLTPPLTSSVARQEQCRQIVLDYKNG 60

QY      61 VTPGITEDQWRAKYVDSAFHPDTGEKVVILGRMSAQVPMNMTITGCMLEFTRKPTV 120
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      61 KVSPTLTVSELWKAKTLVDSTVHPDTGEKMFILGRMSAQMPANMLINGMLSLYRTPGV 120

QY      121 VFWQWVNSFNAINVNSRSGDTPITVRQLGTAVYSATTGAVATAGLKSLLTKHLPPLVG 180
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      121 VFSHWINQSFNAVVTNRSNGSKTSNERLLISYSCATGAMAALSLNAMNKN--SIAA 178

QY      181 RFVPFAVAANANCINIPLMRQRELQVGIPIVADEAGQRLGYSVTAAKQGIQVVISRICMA 240
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      179 RLVPFAVALANTINIPMIRSNEVTEGLELRDENGELLARSRQMAILSIAQVTLISRIAMA 238

QY      241 IPAMAIPLIMDLEK 256
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      239 MPDMVMTPIIMNRITR 254
```

#### RESULT 6

T27337

hypothetical protein Y6E2A.9 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jan-2000

C;Accession: T27337

R;Matthews, L.

submitted to the EMBL Data Library, January 1998

A;Reference number: Z20347

A;Accession: T27337

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-329 <WIL>

A;Cross-references: EMBL:AL021175; PIDN:CAA15970.1; GSPDB:GN00023; CESP:Y6E2A.9

A;Experimental source: clone Y6E2A

C;Genetics:

A;Gene: CESP:Y6E2A.9

A;Map position: 5

A;Introns: 26/1; 99/3; 202/2; 245/1

C;Superfamily: Saccharomyces probable membrane protein YOR271c

Query Match 44.1%; Score 593; DB 2; Length 329;

Best Local Similarity 46.6%; Pred. No. 2.2e-44;

Matches 116; Conservative 39; Mismatches 94; Indels 0; Gaps 0;

```
QY      13 NIOEPRWDQSTFLGRARHFFVTDPNNLLSGAQLASRNIVQNYRAGVTPGITEDQLW 72
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      13 DISKSKMDLDTYSGRVKHYFASANPMTLFTSSNTQEMCRKIVVDYKKGIINPELTMDELW 72

QY      73 RAKYVDSAFHPDTGEKVVILGRMSAQVPMNMTITGCMLEFTRKPTVTVFWQWVNSFNA 132
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      73 SAKILYDSVYHPDTGEKMFELGRMSAQTPANNVITGMLLSCYRTCPGILFSHWINQSFNA 132

QY      133 IVNYSNRSGDTPITVRQLGTAVYSATTGAVATAGLKSLLTKHLPPLVGRVFPAAVAAN 192
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      133 IVNYTNRSNGCRITNQQLLYSYFCATGAATTAALGLNMVKNSHGLAGRLVFAVAVAN 192

QY      193 CINIPLMRQRELQVGIPIVADEAGQRLGYSVTAAKQGIQVVISRICMAIPPLIMD 252
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      193 AINIPMVRANELSEGIELCEDDDHLVAKSKQALALAIQVTLISRLMAMPDVLSPVIMN 252

QY      253 TLEKDFLK 261
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      253 RTRTRATYYK 261
```

#### RESULT 7

T15498

hypothetical protein C14F5.4 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 18-Feb-2000

C;Accession: T15498

R;Minx, P.

submitted to the EMBL Data Library, June 1995

A;Description: The sequence of C. elegans cosmid C14F5.

A;Reference number: Z18361

A;Accession: T15498

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-285 <MIN>

A;Cross-references: EMBL:U29082; NID:g861384; PID:g861388; PIDN:AAA68404.1; CESP:C14F5.4

A;Experimental source: strain Bristol N2

C;Genetics:

A;Gene: CESP:C14F5.4

A;Introns: 31/3; 82/2; 136/2; 227/2

C;Superfamily: hypothetical protein C14F5.4

Query Match 35.9%; Score 483.5; DB 2; Length 285;

Best Local Similarity 44.4%; Pred. No. 7.3e-35;

Matches 103; Conservative 37; Mismatches 89; Indels 3; Gaps 2;

```
QY      16 EPRWDQSTFLGRARHFFVTDPNNLLSGAQLASRNIVQNYRAGVTPGITEDQLWRAK 75
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      40 QPRYDQSTFYGRRLRHPAGMTDPLIAFSSTTELLITASELMQCKREKPPVA-TLEELHRSQ 98

QY      76 YVYDSAFHPDTGEKVVILGRMSAQVPMNMTITGCMLEFTRKPTVTVFWQWVNSFNAIVN 135
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      99 RLCQSAFHPDTGELQNFAGRMCFNVWGCTMLCGAMMIWYKSTPAVIFWQMANQSFNALVN 158

QY      136 YSNRSGDTPITVRQLGTAVYSATTGAVATAGLKS--LTKHLPPLVGRVFPAAVAAN 193
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      159 YTNRAKSTLTTKDLVSYSTAVSGALMAIGLKYTPAKKQSSPLAQRLVPLGAVAVANA 218

QY      194 INIPLMRQRELQVGIPIVADEAGQRLGYSVTAAKQGIQVVISRICMAIPAMA 245
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      219 INIPMRQNELKEGTVTDADGNNVGVSRLAAKAISLVLSRNIIYAPCMS 270
```

#### RESULT 8

S67168

probable membrane protein YOR271c - yeast (*Saccharomyces cerevisiae*)

N;Alternate names: hypothetical protein O5433

C;Species: *Saccharomyces cerevisiae*

C;Date: 12-Jul-1996 #sequence\_revision 12-Jul-1996 #text\_change 19-Apr-2002

C;Accession: S67168; S67173; S72042

R;Jauniaux, J.C.; Poirey, R.

submitted to the Protein Sequence Database, July 1996

A;Reference number: S67143

A;Accession: S67168

A;Molecule type: DNA

A;Residues: 1-327 <JAU>

A;Cross-references: EMBL:Z75179; NID:g1420607; PID:g1420608; MIPS:YOR271c

A;Experimental source: strain S288C

R;Cheret, G.; Sor, F.

submitted to the Protein Sequence Database, July 1996

A;Reference number: S67169

A;Accession: S67173

A;Molecule type: DNA

A;Residues: 1-327 <CHE>

A;Cross-references: EMBL:Z75179; NID:g1420607; PID:g1420608; MIPS:YOR271c

A;Experimental source: strain S288C

R;Cheret, G.; Bernardi, A.; Sor, F.

Yeast 12, 1059-1064, 1996

A;Title: DNA sequence analysis of the VP1-SNF2 region on chromosome XV of *Saccharomyces*

A;Reference number: S72039; MUID:97051594; PMID:8896271

A;Accession: S72042

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA



A/Residues: 1-327 <CHW>  
A/Cross-references: EMBL:X89633; NID:g1279694; PIDN:CAA61777.1; PID:g1279698  
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1995  
C/Genetics:  
A/Cross-references: SGD:S0005797  
A/Map position: 15R  
A/Note: YOR271c  
C/Superfamily: *Saccharomyces* probable membrane protein YOR271c  
C/Keywords: transmembrane protein  
F/184-200/Domain: transmembrane #status predicted <TM1>  
F/276-292/Domain: transmembrane #status predicted <TM2>

Query Match:	34.5%;	Score 464.5;	DB 2;	Length 327;
Best Local Similarity	40.6%;	Pred. No. 4e-33;		
Matches 104;	Conservative 41;	Mismatches 102;	Indels 9;	Gaps 2;

```

QY      8 LPLDINIQEPFWDOSTFLGRARHFFVTDPBNLLLSGAQLEASRNIVONVRAGVTPGIT 67
Db      5 VEPGIDLPESRYDLSTYWGRIRHCAEISDPTMLLTTEKDLAHAREIISAYRHGEJKE--T 62
QY      68 EDOLWRAKYVYDSAFHPDTEKEKVLIGRMSAQVPMNMTTITGCMLTFFYRKPTPTVFWQWVN 127
Db      63 TPEFWRAKQOLDSTVHPDGTGKTVLLPFRRMSNVLSNLVVTWGMTPLGLGTAGTFFWQWAN 122
QY      128 OSFNAIVNYSNRSGDTPITVRQLGTAYVASATTGAVALG-----LKSLTKHLPELVG 180
Db      123 QSLNVAVNSANANKSHPMSTSQLTNTYAATAVASCVALGLNNLVPRLKNISPHSKLITG 182
QY      181 RFVFPFAAVALAANCINIPLMRQRELQVGIPIVADEAGQRLGYSVTAAKOGIFQVVISRICMA 240
Db      183 RLVPFAAVVSAGIVNVFLMRGNRIKGISVFDNSNGDEVGSKKAAFMVAVGETALSRYINA 242
QY      241 IPAMAIPPLIMDTLEK 256
Db      243 TPTWVIPPILVRLQR 258

```

RESULT 9  
T19873  
hypotheical protein C41C4.2 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C/Accession: T19873  
R/Burton, J.  
submitted to the EMBL Data Library, February 1995  
A/Reference number: Z19190  
A/Accession: T19873  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-616 <WIL>  
A/Cross-references: EMBL:Z48045; PIDN:CAA88099.1; GSPDB:GN00020; CESP:C41C4.2  
A/Experimental source: clone C41C4  
C/Genetics:  
A/Gene: CESP:C41C4.2  
A/Map position: 2  
A/Introns: 24/2; 72/3; 214/3; 291/2; 371/2; 419/1; 481/3; 503/1; 530/2

	Query Match	32.2%;	Score 432.5;	DB 2;	Length 616;
	Best Local Similarity	40.0%;	Pred. No. 5.7e-30;		
	Matches 100;	Conservative 43;	Mismatches 96;	Indels 11;	Gaps 6
QY	16 EPRWDSTFLGARHFFVTVDPRNLLSGAQLASRNIVQNRYAGVYTPGITEDQLWRAK	75			
	::                  :   ::  :: ::          :				
Dd	44 EPRFPQDTFLGRYLHCLDVIDPRTLFASNKKLEBSLELNSFKAGTAT-NVPDKSLWEAQ	102			
QY	76 VYVDSA FHPDTGEKVV LIGRMSAQVPMN-MTITGCM LTFYRKPTPVFVFQWVNQSFNATV	134			
	: :                 :   ::				
Dd	103 KLSAILHPDTGEKVLP PPFRRMSGVPEFGMITVTG-MLLPNPSWPTLLFWQMNNQSHNACV	161			
QY	135 NYSNRSGDTPITVRQLGTAYVSATTGAVALALGLKSLTK--HLPP----LVGRFVPPAA	187			
	::    : :         : :				
Dd	162 NYANRNATOPOLSKYTGAYGAAVTAACSISGGLTYFIKASSLPPTTRIIIQRFVPLPA	221			

```

QY      188  VAAANCINIPLMRÖRELÖVGIFPVAD-EAGORLGYSVTAKÖGIFQVVISRICMAIPAMAI 246
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      222  TSLASSLNVICMRWNELETGIQVEKDTGKVGVSKVAAKQAVTDTMTVRAPLVPULLM 281
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      247  PPLIMDTLEK 256
          ||||| : :
Db      282  PECIMPYLER 291
          ||||| : :

```

RESULT 10  
T37847  
probable

transporter - fission yeast (*Schizosaccharomyces pombe*)

C;/Species: *Saccharomycetes* probable  
C;/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jan-2000  
C;/Accession: T37847  
R;/Murphy, L.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Wood, V.  
submitted to the EMBL Data Library, September 1997  
A;/Reference number: Z21749  
A;/Accession: T37847  
A;/Status: preliminary; translated from GB/EMBL/DBJ  
A;/Molecule type: DNA  
A;/Residues: 1-325 <MUR>  
A;/Cross-references: EMBL:Z99162; PIDN:CAB16226.1; GSPDB:GNU00066; SPDB:SPAC17G6.15c  
A;/Experimental source: strain 972h-; cosmid c17G6  
C;/Genetics:  
A;/Gene: SPDB:SPAC17G6.15c  
A;/Map position: 1  
A;/Introns: 33/2; 70/1; 94/3; 206/3  
C;/Superfamily: *Saccharomycetes* probable membrane protein YOR271c

Query Match	30.6%;	Score 411.5;	DB 2;	Length 325;
Best Local Similarity	36.3%;	Pred. No. 1.8e-28;		
Matches 89;	Conservative 48;	Mismatches 99;	Indels 9;	Gaps 2;

QY		24	FLGRARHEFTVTDPRLNLLSGAOLEASRNIVONYRAGVVTPTGITEDOLWRAKYVYDSAFH	83
Dd		19	FMLMRQHAMDITDPRTLTSTQDLNSAVKTELEDYGAKIAQ--LDETVMHAKKIVDSTLIH	76
QY		84	PDTGEKVVLIGRMSAQVPMMNTTTCMLTFYRKPTPVFWQMWNQSFNALVNYSNRSGDT	143
Dd		77	PDKPEPVELPERMSCFVLNLNVTAQNLOPNLGTAGTVFMQWMNQSVNVAFAFSANANKST	136
QY		144	PI TVRQLGTAYVSATTTGA VATALGLSLTKHL-----PPLVGRFVPPAAVVAANCINI	196
Dd		137	QLTLPQMTKSXYIYAVSASCVAIGLNKTIVRMNFLLSSSKAVLGRLTFFA AVASAGVLNV	196
QY		197	PLMRORELQVGIPVADAGQRIGYSVTA AKOGIFQVVISRICMAIPAMAIPP LMDTLEK	256
Dd		197	FLMRGEELRQGIDVPDKEGESLGKSKAAFYAVGETALS RVINASPIMVIPPLVLMRLQK	256
QY		257	KDFLK 261	
Dd		257	QNMILR 261	

RESULT 11  
T02841

mitochondrial tricarboxylate carrier MTCC [imported] - Leishmania major (strain Friedlir  
C/Species: Leishmania major  
C/Date: 24-Mar-1999 #sequence revision 24-Mar-1999 #text change 19-May-2000

C/Accession: E81461; T02841  
R/Myler, P.J.; Audleman, L.; devos, T.; Hixson, G.; Kiser, P.; Lemley, C.; Magness, C.;  
Proc. Natl. Acad. Sci. U.S.A. 96, 2902-2906, 1999  
A/Title: Leishmania major Friedlin chromosome 1 has an unusual distribution of protein-c  
A/Reference number: A81455; NUID:99178987; PMID:10077609

A/Accession: E81461  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-326 <PYL>  
A/Cross-references: GB:AE001274; NID:g3264850; PIDN:AAC24664.1; PID:g1617564; GSPDB:GNOC  
A/Experimental source: strain MHOM/IL/81/Friedlin  
C/Genetics:  
A/Gene: MTCC





QY 96 ---MSAQPMMNTITGCMLEFYRKPTV--VFQWV-----NOSF 130  
Db 98 YRDMSL---NLTKGVSAKAFQTEELRLKLPKMGELNDGEAGYANPSSGMANAQSV 154  
QY 131 NAIVNYSNRSGDPTITVROLGTAYVSATGAVATAGLKSITKHLPLVGRFVP---FAA 187  
Db 155 KSVVNY-----LAHAGVSFISGPEGIVEELITBENVVKGVRTTGAYMAKLIFA 204  
QY 188 VAAANCINIPLMRQRELQVGPVA 211  
Db 205 TGAWTASLLPNDHTRFLATGQFVA 228

RESULT 15

156551  
neurotrimin - rat  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 26-Jul-1996 #sequence\_revision 26-Jul-1996 #text\_change 19-May-2000  
C/Accession: 156551  
R/Struyk, A.F.; Canoll, P.D.; Wolfgang, M.J.; Rosen, C.L.; D'Eustachio, P.; Salzer, J.L.  
J. Neurosci. 15, 2141-2156, 1995  
A/Title: Cloning of neurotrimin defines a new subfamily of differentially expressed neur  
A/Reference number: 156551; MID:95198094; PMID:7891157  
A/Accession: 156551  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-344 <RES>  
A/Cross-references: EMBL:U16845; NID:g755184; PIDN:AAA67445.1; PID:g755185  
C/Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-termin

Query Match 6.7%; Score 89.5; DB 2; Length 344;

Best Local Similarity 24.8%; Pred.No.3.4;  
Matches 52; Conservative 28; Mismatches 95; Indels 35; Gaps 10;

QY 2 ESKMGEPLDINIOEPRW-DQSTFLGRARHFTVTDPRNLLSGAQLASRNI--VQNYR 58  
Db 51 ESATLRCTIDNRVTRVAVLNRSTILYAGNDKWCL-DPRVLLSNTQTQYSIEIQNVDDYD 109  
QY 59 AGVTPGITEDQLWRAKCVYDSAFHPDTGEKVVLIGRMSAQV-----PMNMTIT 107  
Db 110 EGPYTCVQTDN-----HPKT-SRVHLIVQVSPKIVEISSDISINEGNNISLT 156  
QY 108 GCMLEFYRKPTPTVVFQWVNSQFNAIVNYSNRSGDPTITVROLGTAYVSATGAVATAG 167  
Db 157 -CIAT-GRPEPTVT-WRHISPKAVGVSEDEYLEIQGITREQSGEYECASANDVAAAPVVR 213  
QY 168 LKSLTKHLPLV---GRFVPFAAANAANC 193  
Db 214 RVNVTVNPYPYISEAKGTGVFGQKTLQC 243

Search completed: February 3, 2004, 09:13:48  
Job time : 22 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 3, 2004, 09:11:53 ; Search time 17 Seconds  
(without alignments)  
721.998 Million cell updates/sec

Title: US-09-990-415A-2  
Perfect score: 1345  
Sequence: 1 MESKMGELPLDINIQEPRWD.....PAMAIPLIMDTLEKKDFLK 261

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues  
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_41:\*

Pred. NO. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1326	98.6	321	1 SFX3_HUMAN	Q9bwm7 homo sapien
2	1272	94.6	321	1 SFX3_RAT	Q9jhy2 rattus norv
3	1269	94.3	321	1 SFX3_MOUSE	Q91v61 mus musculu
4	1089	81.0	322	1 SFX1_HUMAN	Q9h9b4 homo sapien
5	1064	79.1	322	1 SFX1_RAT	Q63965 rattus norv
6	1059	78.7	322	1 SFX1_MOUSE	Q99jrl mus musculu
7	753	56.0	322	1 SFX2_MOUSE	Q925n2 mus musculu
8	749	55.7	322	1 SFX2_HUMAN	Q96nb2 homo sapien
9	639	47.5	329	1 YP22_CAEBL	Q09201 caenorhabdi
10	484.5	36.0	340	1 SFX5_HUMAN	Q8td22 homo sapien
11	484.5	36.0	342	1 SFX5_MOUSE	Q925n0 mus musculu
12	432.5	32.2	600	1 SRE2_CAEBL	Q09273 caenorhabdi
13	113	8.4	313	1 SFX4_MOUSE	Q925n1 mus musculu
14	96	7.1	836	1 GCSR_HUMAN	Q99062 homo sapien
15	89.5	6.7	344	1 NTRI_RAT	Q62718 rattus norv
16	88.5	6.6	344	1 NTRI_HUMAN	Q9p121 homo sapien
17	88.5	6.6	344	1 NTRI_MOUSE	Q99p10 mus musculu
18	88	6.5	594	1 RCO3_NEUCR	Q92253 neurospora
19	87.5	6.5	234	1 YE99_ARCFU	Q28773 archaeoglob
20	85	6.3	914	1 PERT_RAT	P14650 rattus norv
21	85	6.3	926	1 PERT_PIG	P09933 sus scrofa
22	84	6.2	272	1 EUTC_PSEPK	Q88gf2 pseudomonas
23	82.5	6.1	274	1 PTRD_KLEPN	P37083 klebsiella
24	82	6.1	146	1 HBB1_MOUSE	P02088 mus musculu
25	82	6.1	836	1 VG26_BPMUS	Q05233 mycobacteri
26	81.5	6.1	531	1 UD16_MOUSE	Q64435 mus musculu
27	81	6.0	519	1 NIFL_AZOVI	P30663 azotobacter
28	81	6.0	703	1 CDGT_BACS2	P31746 bacillus sp
29	81	6.0	914	1 PERT_MOUSE	P35419 mus musculu
30	81	6.0	2505	1 FAS_RAT	P12785 rattus norv
31	80.5	6.0	387	1 MURG_ZYMMO	Q9rm6 zymomonas m
32	80.5	6.0	393	1 XYLH_ECOLI	P37389 escherichia
33	80.5	6.0	913	1 DPOL_CHV2	P30320 chlorella v

34	80	5.9	501	1 DLDH_PEA	P31023 pisum sativ
35	80	5.9	617	1 PYG1_PSEAE	Q06583 pseudomonas
36	80	5.9	688	1 PYG2_PSEAE	Q06584 pseudomonas
37	80	5.9	815	1 KINH_CAEBL	P34540 caenorhabdi
38	79.5	5.9	353	1 CEPV_CHICK	Q90773 gallus gall
39	79.5	5.9	529	1 UD16_RAT	Q90773 gallus gall
40	79.5	5.9	693	1 GUAA_HUMAN	P49915 homo sapien
41	79	5.9	337	1 G55A_CHICK	Q98892 gallus gall
42	79	5.9	455	1 YWAD_BACSU	P25152 bacillus su
43	78.5	5.8	412	1 YADC_ECOLI	P31058 escherichia
44	78.5	5.8	470	1 YUJR_ECOLI	P39389 escherichia
45	78	5.8	4349	1 PAT2_HUMAN	Q9nyq8 homo sapien

## ALIGNMENTS

```
RESULT 1
ID      SFX3_HUMAN          STANDARD;          PRT;          321 AA.
AC      Q9BWM7; Q9NTP4;
DT      28-FEB-2003 (Rel. 41, Created)
DT      28-FEB-2003 (Rel. 41, Last sequence update)
DT      15-SEP-2003 (Rel. 42, Last annotation update)
DE      Sideroflexin 3.
GN      SFXN3.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxId=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Heath P.;
RL      Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN      [2]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Eye;
RX      MEDLINE=22388257; PubMed=12477932;
RA      Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA      Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA      Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA      Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA      Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA      Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA      Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA      Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA      Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA      Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA      Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA      Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA      Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA      Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA      Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA      Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA      Schnerch A., Schein J.B., Jones S.J.M., Marra M.A.;
RT      "Generation and initial analysis of more than 15,000 full-length
RL      human and mouse cDNA sequences.";
RN      Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN      [3]
RP      SUBCELLULAR LOCATION.
RA      Rabiloud T.;
RL      Unpublished observations (FEB-2002).
CC      -1- FUNCTION: Potential iron transporter.
CC      -1- SUBCELLULAR LOCATION: Mitochondrial.
CC      -1- SIMILARITY: Belongs to the sideroflexin family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.1sb-sib.ch/announce/
CC      or send an email to license@1sb-sib.ch).
```

```

CC -----
DR EMBL; AL133215; -, NOT ANNOTATED_CDS.
DR EMBL; BC000124; AAH00124.1; -.
DR InterPro; IPR004686; Mtc.
DR Pfam; PF03820; Mtc; 1.
DR ProDom; PD006986; Mtc; 1.
DR TIGRFAMs; TIGR00798; mtc; 1.
KW Transport; Iron transport; Iron; Mitochondrion; Transmembrane.
FT TRANSMEM 146 164 POTENTIAL.
FT TRANSMEM 174 194 POTENTIAL.
FT TRANSMEM 226 246 POTENTIAL.
FT TRANSMEM 266 286 POTENTIAL.
FT CONFLICT 112 112 K -> Q (IN REF. 1).
SQ SEQUENCE 321 AA; 35503 MW; 54718C600B3D34BD CRC64;

Query Match      98.6%; Score 1326; DB 1; Length 321;
Best Local Similarity 100.0%; Pred. No. 4.5e-110;
Matches 257; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 MGELPLDINIQEPRWDQSTFLGRARHFTVTDPRLLLSGAQLASRNIVQNYRAGVTP 64
   |||||
DB 1 MGELPLDINIQEPRWDQSTFLGRARHFTVTDPRLLLSGAQLASRNIVQNYRAGVTP 60

QY 65 GITEDQLMRAKYVYDSAFHPDTEGEKVVIGRMSAQVPMNMTITGCMLTFRKPTPVFWQ 124
   |||||
DB 61 GITEDQLMRAKYVYDSAFHPDTEGEKVVIGRMSAQVPMNMTITGCMLTFRKPTPVFWQ 120

QY 125 WVNQSFNAIVNYSNRSGDPTITVRQLGTAYVSATGAVATAGLKSITKHLPLVGRFVP 184
   |||||
DB 121 WVNQSFNAIVNYSNRSGDPTITVRQLGTAYVSATGAVATAGLKSITKHLPLVGRFVP 180

QY 185 FFAVAAANCINIPLMRQRELQVGIPVADGAGRLGYSVTAAKGIFQVVISRICMAIPAM 244
   |||||
DB 181 FFAVAAANCINIPLMRQRELQVGIPVADGAGRLGYSVTAAKGIFQVVISRICMAIPAM 240

QY 245 AIPPLIMDTLEKKDFLK 261
   |||||
DB 241 AIPPLIMDTLEKKDFLK 257

RESULT 2
SFX3_RAT      STANDARD;      PRT;      321 AA.
ID SFX3_RAT
AC Q9JHY2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Sideroflexin 3.
GN SFXN3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RA Mashima H., Kojima I.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
CC -|- FUNCTION: Potential iron transporter.
CC -|- SUBCELLULAR LOCATION: Mitochondrial (By similarity).
CC -|- SIMILARITY: Belongs to the sideroflexin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF276997; AAF78249.1; -.
DR InterPro; IPR004686; Mtc.
RA Pfam; PF03820; Mtc; 1.

```

```

DR ProDom; PD006986; Mtc; 1.
KW TIGRFAMs; TIGR00798; mtc; 1.
DR Transport; Iron transport; Iron; Mitochondrion; Transmembrane.
FT TRANSMEM 146 164 POTENTIAL.
FT TRANSMEM 174 194 POTENTIAL.
FT TRANSMEM 226 246 POTENTIAL.
FT TRANSMEM 266 286 POTENTIAL.
SQ SEQUENCE 321 AA; 35433 MW; 61D9DE679A7E6124 CRC64;

Query Match      94.6%; Score 1272; DB 1; Length 321;
Best Local Similarity 95.3%; Pred. No. 2.7e-105;
Matches 245; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

QY 5 MGELPLDINIQEPRWDQSTFLGRARHFTVTDPRLLLSGAQLASRNIVQNYRAGVTP 64
   |||||
DB 1 MGELPLDINIQEPRWDQSTFLGRARHFTVTDPRLLLSGAQLASRNIVQNYRAGVTP 60

QY 65 GITEDQLMRAKYVYDSAFHPDTEGEKVVIGRMSAQVPMNMTITGCMLTFRKPTPVFWQ 124
   |||||
DB 61 GITEDQLMRAKYVYDSAFHPDTEGEKVVIGRMSAQVPMNMTITGCMLTFRKPTPVFWQ 120

QY 125 WVNQSFNAIVNYSNRSGDPTITVRQLGTAYVSATGAVATAGLKSITKHLPLVGRFVP 184
   |||||
DB 121 WVNQSFNAIVNYSNRSGDPTITVRQLGTAYVSATGAVATAGLKSITKHLPLVGRFVP 180

QY 185 FFAVAAANCINIPLMRQRELQVGIPVADGAGRLGYSVTAAKGIFQVVISRICMAIPAM 244
   |||||
DB 181 FFAVAAANCINIPLMRQRELQVGIPVADGAGRLGYSVTAAKGIFQVVISRICMAIPAM 240

QY 245 AIPPLIMDTLEKKDFLK 261
   |||||
DB 241 AIPPLIMDTLEKKDFLK 257

RESULT 3
SFX3_MOUSE    STANDARD;      PRT;      321 AA.
ID SFX3_MOUSE
AC Q91V61; Q8C1Z2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Sideroflexin 3.
GN SFXN3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1), AND TISSUE SPECIFICITY.
RX MEDLINE=21172735; PubMed=11274051;
RA Fleming M.D., Campagna D.R., Haslett J.N., Trenor C.C. III,
RA Andrews N.C.;
RT "A mutation in a mitochondrial transmembrane protein is responsible
RT for the pleiotropic hematological and skeletal phenotype of
RT flexed-tail (f/f) mice."
RT Genes Dev. 15:652-657(2001).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RX MEDLINE=22354683; PubMed=1246851;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chochia C., Corbani L.B., Cousins S.,
RA Daller E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawai H., Kawasaki Y., Kedzierski R.M., King P.A.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,

```



RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
RA Sandelin A., Schneider C., Sempé C.A., Setou M., Shimada K.,  
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
RA Hara A., Haeizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
RA Birney E., Hayashizaki Y.,  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs.";  
RL Nature 420:563-573 (2002).  
[3]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RC TISSUE=Breast;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Bueltow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Buterfield V.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
CC -1- FUNCTION: Potential iron transporter.  
CC -1- SUBCELLULAR LOCATION: Mitochondrial (By similarity).  
CC -1- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=1;  
CC IsoId=Q91V61-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=Q91V61-2; Sequence=VSP\_007388;  
CC Note=No experimental confirmation available;  
CC -1- TISSUE SPECIFICITY: Widely expressed.  
CC -1- SIMILARITY: Belongs to the sideroflexin family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF325262; AAK39430.1; -;  
DR EMBL; AK089985; BAC41029.1; -;  
DR EMBL; BC012208; AAH12208.1; -;  
DR MGD; MGI:2137679; Sfxn3.  
DR InterPro; IPR004686; Mtc.  
DR Pfam; PF03820; Mtc; 1.  
DR ProDom; PD006986; Mtc; 1.  
DR TIGRFAMs; TIGR00798; mtc; 1.  
KW Transport; Iron transport; Iron; Mitochondrion; Transmembrane;  
KW Alternative splicing.  
FT TRANSMEM 146 164 POTENTIAL.  
FT TRANSMEM 174 194 POTENTIAL.  
FT TRANSMEM 225 245 POTENTIAL.  
FT TRANSMEM 266 286 POTENTIAL.

FT VARSPPLIC 112 144 Missing (in isoform 2).  
FT FT CONFLICT 243 243 /FTId=VSP\_007388.  
SQ SEQUENCE 321 AA; 35406 MW; 9B9816A54B23F4BC CRC64;  
Query Match 94.3%; Score 1269; DB 1; Length 321;  
Best Local Similarity 95.3%; Pred. No. 5e-105;  
Matches 245; Conservative 7; Mismatches 5; Indels 0; Gaps 0;  
QY 5 MGELPLDINIOEPRWDSFELGRARHFTVTDPRNLLSGAQLBASRNIVQNYRAGVTP 64  
Db 1 MGDLPLNINIOEPRWDSFELGRARHFTVTDPRNLLSGEQLBASRNIVQNYRAGVATP 60  
QY 65 GITEDQLMRAKYVYDSAFHPDTGKLVILGRMSAQVPMNMTITGCMITFYRKTPTVVFVWQ 124  
Db 61 GITEDQLMRAKYVYDSAFHPDTGKLVILGRMSAQVPMNMTITGCMITFYRKTPTVVFVWQ 120  
QY 125 WYNQSFNAIVNYSNRSGDPTITVRQLGTAVYSATGAVATAGLSLTKHLPLVGRFVP 184  
Db 121 WYNQSFNAIVNYSNRSGDAPITVQQLGTAVYSATGAVATAGLSLTKHLPLVGRFVP 180  
QY 185 PAAVAAANCINIPLMRQRELOVGI PVADAGQRGLGYSVTAAKOGIFQVVISRICMAIPAM 244  
Db 181 PAAVAAANCINIPLMRQRELOVGI PVTDAGQRGLGHSVTAAKOGIFQVVISRICMAIPAM 240  
QY 245 AIPPLIMDTLEKDFLK 261  
Db 241 AIPVIMNTLEKDFLK 257  
RESULT 4  
SFX1\_HUMAN STANDARD; PRT; 322 AA.  
ID Q9H9B4; Q9H9B3;  
AC Q9H9B4; Q9H9B3; (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Sideroflexin 1.  
GN SFXN1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RL Li N., Chen T., Wan T., Zhang W., Cao X.;  
RN Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Mammary gland, and Teratocarcinoma;  
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,  
RA Wagatsuma M., Hosokiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,  
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,  
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y.,  
RA Ninomiya K., Iwayanagi T.;  
RT "NEDO human cDNA sequencing project.";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 19-322 FROM N.A.  
RC TISSUE=Lung;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Bueltow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

```

RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzyzinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -1- FUNCTION: Might be involved in the transport of a component
CC required for iron utilization into or out of the mitochondria.
CC -1- SUBCELLULAR LOCATION: Mitochondrial (By similarity).
CC -1- SIMILARITY: Belongs to the sideroflexin family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AF327346; AAL56007.1; -
DR EMBL; AK022287; BAB14003.1; ALT_SEQ.
DR EMBL; AK022938; BAB14318.1; ALT_INIT.
DR EMBL; BC020517; AAH20517.1; ALT_INIT.
DR Genew; HGNC:16085; SFXN1.
DR InterPro; IPR004686; Mtc.
DR Pfam; PF03820; Mtc; 1.
DR ProDom; PD006986; Mtc; 1.
DR TIGRFAMs; TIGR00798; mtc; 1.
KW Transport; Iron transport; Iron; Mitochondrion; Transmembrane.
FT TRANSMEM 103 120 POTENTIAL.
FT TRANSMEM 147 167 POTENTIAL.
FT TRANSMEM 175 195 POTENTIAL.
FT TRANSMEM 229 249 POTENTIAL.
FT TRANSMEM 267 287 POTENTIAL.
FT CONFLICT 1 63 MSGELPNINIKPRWDQSTFIGRANHFVTVDPRNILLTN
FT CONFLICT 73 73 BAB14003).
FT CONFLICT 291 322 SMSVTSLEAEIQAKIQESHPELRVYFNKGL -> RYVEV
FT CONFLICT 322 322 IRONHEV (IN REF. 2; BAB14003).
SQ SEQUENCE 322 AA; 35619 MW; 47E03172F27990DB CRC64;

Query Match 81.0%; Score 1089; DB 1; Length 322;
Best Local Similarity 78.5%; Pred. No. 4.3e-89;
Matches 201; Conservative 28; Mismatches 27; Indels 0; Gaps 0;

QY 6 GELPLDINIOEPRWDQSTFIGRANHFVTVDPRNILLSGAQLASRNIVQNYRAGVTPG 65
DB 3 GELPNINIKPRWDQSTFIGRANHFVTVDPRNILLTNEQLSARKIVHDYRQGIIVPG 62
QY 66 ITEDQLMRKAYVYDSAFHPDTGKLVIGRMSAQVPMNMNTITGCMLEFYRKPTVVFQW 125
DB 63 LTENELMRKAYVYDSAFHPDTGKMLIGRMSAQVPMNMNTITGCMTFYRTPAVLFWQW 122
QY 126 VNQSFNAIVNYSNRSGDPTITVRQLGTAVYSATTGAATATGLKSLTKHLPLVGRFVPF 185
DB 123 INQSFNAVNYNTRSGDAPLTVNLGTAVYSATTGAATATGLNALTGHVSPLIGRFVPF 182
QY 186 AAVAAANCINIPLMRQRELQVGIPIVADAGORLGYSVTAAGKQIFQVVISRICMAIPAMA 245
DB 183 AAVAAANCINIPLMRQRELKVGIPVTDENGNGRLGESANAAKAQTQVVISRILMAAPGMA 242
QY 246 IPPLIMDTLEKDFLK 261
DB 243 IPPFIMNTLEKKAFLK 258

```

```

ID SFX1_RAT STANDARD; PRT; 322 AA.
AC Q63965;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Sideroflexin 1 (Tricarboxylate carrier protein).
GN SFXN1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=94179133; PubMed=8132491;
RA Azzi A., Glerum M., Koller R., Mertens W., Spycher S.;
RT "The mitochondrial tricarboxylate carrier.";
RL J. Bioenerg. Biomembr. 25:515-524(1993).
CC -1- FUNCTION: Might be involved in the transport of a component
CC required for iron utilization into or out of the mitochondria.
CC -1- SUBCELLULAR LOCATION: Mitochondrial.
CC -1- TISSUE SPECIFICITY: Expressed in liver and brain.
CC -1- SIMILARITY: Belongs to the sideroflexin family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; S70011; AAB30258.1; ALT_INIT.
DR InterPro; IPR004686; Mtc.
DR Pfam; PF03820; Mtc; 1.
DR ProDom; PD006986; Mtc; 1.
DR TIGRFAMs; TIGR00798; mtc; 1.
KW Transport; Iron transport; Iron; Mitochondrion; Transmembrane.
FT TRANSMEM 103 120 POTENTIAL.
FT TRANSMEM 147 167 POTENTIAL.
FT TRANSMEM 175 195 POTENTIAL.
FT TRANSMEM 229 249 POTENTIAL.
FT TRANSMEM 267 287 POTENTIAL.
SQ SEQUENCE 322 AA; 35546 MW; D28CBD898E8ABC5D CRC64;

Query Match 79.1%; Score 1064; DB 1; Length 322;
Best Local Similarity 77.3%; Pred. No. 7.1e-87;
Matches 198; Conservative 27; Mismatches 31; Indels 0; Gaps 0;

QY 6 GELPLDINIOEPRWDQSTFIGRANHFVTVDPRNILLSGAQLASRNIVQNYRAGVTPG 65
DB 3 GELPNINIKPRWDQSTFIGRASHFFVTVDKNIILLTNEQLNARKVHDYRQGIIVPG 62
QY 66 ITEDQLMRKAYVYDSAFHPDTGKLVIGRMSAQVPMNMNTITGCMLEFYRKPTVVFQW 125
DB 63 LTENELMRKAYVYDSAFHPDTGKMTLIGRMSAQVPMNMNTITGCMTFYRTPAVLFWQW 122
QY 126 VNQSFNAIVNYSNRSGDPTITVRQLGTAVYSATTGAATATGLKSLTKHLPLVGRFVPF 185
DB 123 INQSFNAVNYNTRSGDAPLTVNLGTAVYSATTGAATATGLNALTGHVSPLIGRFVPF 182
QY 186 AAVAAANCINIPLMRQRELQVGIPIVADAGORLGYSVTAAGKQIFQVVISRICMAIPAMA 245
DB 183 AAVAAANCINIPLMRQRELKVGIPVTDENGNGRLGESANAAKAQTQVVISRILMAAPGMA 242
QY 246 IPPLIMDTLEKDFLK 261
DB 243 IPPFIMNTLEKKAFLK 258

```

RESULT 5  
SFX1\_RAT

RESULT 6  
SFX1\_MOUSE  
ID SFX1\_MOUSE STANDARD; PRT; 322 AA.

AC Q99UR1; Q9CZG4;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Sideroflexin 1.  
GN SFXN1 OR F.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND DISEASE.  
RX MEDLINE=21172735; PubMed=11274051;  
RA Fleming M.D., Campagna D.R., Haslett J.N., Trenor C.C. III,  
RA Andrews N.C.;  
RT "A mutation in a mitochondrial transmembrane protein is responsible  
RT for the pleiotropic hematological and skeletal phenotype of  
RT flexed-tail (f/f) mice.";  
RL Genes Dev. 15:652-657(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX STRAIN=C57BL/6J; TISSUE=Embryo;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyo-oka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX TISSUE=Breast;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Uebli T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
CC -1- FUNCTION: Might be involved in the transport of a component  
CC required for iron utilization into or out of the mitochondria.  
CC -1- SUBCELLULAR LOCATION: Mitochondrial.  
CC -1- TISSUE SPECIFICITY: Widely expressed, with highest expression in  
CC kidney and liver.  
CC -1- DEVELOPMENTAL STAGE: Very high levels in the liver during the

CC period of embryonic hepatic hemopoiesis.  
CC -1- DISEASE: Defects in SFXN1 are the cause of a transitory  
CC hypochromic, microcytic anemia characterized by a large number of  
CC siderocytes containing non-heme iron granules. The anemia begins  
CC at 12 dpc, is most intense at 15 dpc and is still severe at birth,  
CC but disappears by 2 weeks of age. Mutant adults are no longer  
CC anemic, but they have an impaired response to hemopoietic stress.  
CC Most homozygotes also have an flexed tails and a belly spot.  
CC -1- SIMILARITY: Belongs to the sideroflexin family.  
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A  
CC FRAMESHIFT IN POSITION 282.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF325260; AAK39428.1; -;  
DR EMBL; AK012650; -; NOT ANNOTATED\_CDS.  
DR EMBL; BC005743; AA05743.1; -;  
DR MGD; MGI:2137677; Sfxn1.  
DR GO; GO:0005739; Cytochondrion; IDA.  
DR GO; GO:0030218; Erythrocyte differentiation; IMP.  
DR GO; GO:0006826; P:iron ion transport; IMP.  
DR InterPro; IPR004686; Mtc.  
DR Pfam; PF03820; Mtc; 1.  
DR ProDom; PD006986; Mtc; 1.  
DR TIGRFAMs; TIGR00798; mtc; 1.  
KW Transport; Iron transport; Iron; Mitochondrion; Transmembrane.  
FT TRANSMEM 103 120 POTENTIAL.  
FT TRANSMEM 147 167 POTENTIAL.  
FT TRANSMEM 175 195 POTENTIAL.  
FT TRANSMEM 229 249 POTENTIAL.  
FT TRANSMEM 267 287 POTENTIAL.  
SQ SEQUENCE 322 AA; 35649 MW; E3B055CB03CEDFA7 CRC64;  
  
Query Match 78.7%; Score 1059; DB 1; Length 322;  
Best Local Similarity 77.3%; Pred. No. 2e-86;  
Matches 198; Conservative 26; Mismatches 32; Indels 0; Gaps 0;  
  
QY 6 GELPLDINIQEPRWDQSTFLGRARHFTVTDPRNLLISGAQLASRNIVQNYRAGVTPG 65  
DB 3 GEVPPNINKEPRWDQSTFIGRASHFTVTDPRNILLTNEQLENARKVVDYRQGIYPAG 62  
QY 66 ITBDQLRAKYVYDSAFHPDTEKVVILGRMSAQVPMNTITGCMLTFRKPTVFWQW 125  
DB 63 LTENELMRKAYVDSAFHPDTEKVVILGRMSAQVPMNTITGCMLTFRKPTVFWQW 122  
QY 126 VNQSFNAIVNYSNRSGDPTITVRQLGTAVYSATTGAVATAGLSLTKHLPLVGRFVPF 185  
DB 123 INQSFNAVNYNTRSGDAPLTVNEIGTAVYSATTGAVATAGLSLTKRVSPPIGRFVPF 182  
QY 186 AAVAANNCINIPLMRQRELQVGPVADAGQRLGYSVTAAGQIFQVVISRICMAIPAMA 245  
DB 183 AAVAANNCINIPLMRQRELKVGIPYTDENGTRLGESTNAKQALTVVVISRIIMAAPGMA 242  
QY 246 IPPLINDTLKKDFLK 261  
DB 243 IPPFINMTLEKKAFLK 258  
  
RESULT 7  
SFX2\_MOUSE  
ID\_SFX2\_MOUSE STANDARD; PRT; 322 AA.  
AC Q925N2;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Sideroflexin 2.  
GN SFXN2.



```
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
NC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RX MEDLINE=21172735; PubMed=11274051;
RA Fleming M.D., Campagna D.R., Haslett J.N., Trenor C.C. III,
RA Andrews N.C.;
RT "A mutation in a mitochondrial transmembrane protein is responsible
RT for the pleiotropic hematological and skeletal phenotype of
RT flexed-tail (F/f) mice.";
RRL Genes Dev. 15:652-657(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=kidney;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshlyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.U., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulys S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schultz J., Myers R.M.,
RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schermer A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RLL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -1- FUNCTION: Potential iron transporter.
CC -1- SUBCELLULAR LOCATION: Mitochondrial (By similarity).
CC -1- TISSUE SPECIFICITY: Expressed in brain, heart, kidney, spleen,
CC thymus, liver, stomach and skin.
CC -1- SIMILARITY: Belongs to the sideroflexin family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outpost at
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AF325261; AAK39429.1; -.
DR EMBL; BC019808; AAH19808.1; -.
DR MGI; MGI:2137678; Sfmx2.
DR InterPro; IPR004686; Mtc.
DR Pfam; PF03820; Mtc_1.
DR ProDom; PD006986; Mtc_1.
DR TIGRFAMs; TIGR00798; mtc_1.
KW Transport; Iron transport; Iron; Mitochondrion; Transmembrane.
FT TRANSMEM 99 119 POTENTIAL.
FT TRANSMEM 147 167 POTENTIAL.
FT TRANSMEM 174 194 POTENTIAL.
FT TRANSMEM 223 243 POTENTIAL.
FT TRANSMEM 266 286 POTENTIAL.
FT CONFLICT 3 3 G -> A (IN REF. 2).
SQ SEQUENCE 322 AA; 36141 MW; 00D623689BBB983C CRC64;
```

[illegible]



CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; BC022091; AAH22091.1; -  
DR EMBL; AK055711; BAB70993.1; ALT\_INIT.  
DR Genew; HGNC:16086; SFXN2.  
DR InterPro; IPR004686; Mtc.  
DR Pfam; PF03820; Mtc; 1.  
DR ProDom; PD006986; Mtc; 1.  
DR TIGRFAMs; TIGR00798; mtc; 1.  
KW Transport; Iron transport; Iron; Mitochondrion; Transmembrane.  
FT TRANSMEM 100 122 POTENTIAL.  
FT TRANSMEM 142 164 POTENTIAL.  
FT TRANSMEM 174 192 POTENTIAL.  
FT TRANSMEM 228 250 POTENTIAL.  
FT TRANSMEM 265 287 POTENTIAL.  
FT CONFLICT 145 204 OMALSYFATTTTAVATAVGMNMLTKKAPPLVGRWVPFAAVA  
AANCVNIPMRQDELIGI -> KRPPWMAAGCPLPLMLRL  
TVSISP (IN REF. 2).  
SQ SEQUENCE 322 AA; 36232 MW; 329D577B2583DF31 CRC64;

Query Match 55.7%; Score 749; DB 1; Length 322;  
Best Local Similarity 58.6%; Pred. No. 5.4e-59;  
Matches 146; Conservative 38; Mismatches 65; Indels 0; Gaps 0;

QY 13 NIOEPRWDQSTFLGRARHFTVTDPRNLLLSGAQLASRNIVQYRAGVTPGITEDQLW 72  
Db 9 NIDAPRWDRTEFLGRVGHFLNITDPRTFVSERELDMAKVWEKSRMGVPPGTQVEQLL 68  
QY 73 RAKVYDSEAFHPDTEGEKVVLLIGRMSAQPVMNMTITGCMLEFYRKTPTVVFQWVNOSFNA 132  
Db 69 YAKKLYDSAFHPDTEGEKVNIVGRMSRQLPGCMITGFMLOFYRTMPAVIFWQWVNOSFNA 128  
QY 133 IVNYSNRSQDPTITVRQLGTAVYSATGAVATAGLKSLLTKHLPLVGRFVPFAVAANA 192  
Db 129 LVNVTNRNNAASPTSVRQMALSYFATTTAVATAVGMNMLTKKAPPLVGRWVPFAVAANA 188  
QY 193 CINIPLMRQRELQVIGIPVADAGRLGYSVTAAGKQIFQVVISRICMAIPMAIPLIMD 252  
Db 189 CVNIPMRQDELIGICVKDRNENIGHSRRAAIGITQVVISRITWSAPGMILLPVIME 248  
QY 253 TLEKKDFLK 261  
Db 249 RLEKLFHMQ 257

RESULT 9  
YP22\_CAEEL STANDARD; PRT; 329 AA.  
ID YP22\_CAEEL  
AC Q09201;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Hypothetical protein AH6.2 in chromosome II.  
AH6.2.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Jassal B.;  
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC -!- SIMILARITY: Belongs to the sideroflexin family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; 248009; CAA88076.1; -  
DR PIR; T18612; T18612.  
DR WormPep; AH6.2; CE01456.  
DR InterPro; IPR004686; Mtc.  
DR Pfam; PF03820; Mtc; 1.  
DR ProDom; PD006986; Mtc; 1.  
DR TIGRFAMs; TIGR00798; mtc; 1.  
KW Hypothetical protein; Transport; Transmembrane.  
FT TRANSMEM 150 168 POTENTIAL.  
FT TRANSMEM 178 198 POTENTIAL.  
FT TRANSMEM 274 294 POTENTIAL.  
SQ SEQUENCE 329 AA; 36755 MW; 0F2D38989D438288 CRC64;

Query Match 47.5%; Score 639; DB 1; Length 329;  
Best Local Similarity 49.0%; Pred. No. 3.1e-49;  
Matches 122; Conservative 44; Mismatches 83; Indels 0; Gaps 0;

QY 13 NIOEPRWDQSTFLGRARHFTVTDPRNLLLSGAQLASRNIVQYRAGVTPGITEDQLW 72  
Db 13 DISKPRWDQSTFYAGRAKHFFSSTNPLTFSSRIQOEKCKEITNVTGVISPTLTVDELW 72  
QY 73 RAKVYDSEAFHPDTEGEKVVLLIGRMSAQPVMNMTITGCMLEFYRKTPTVVFQWVNOSFNA 132  
Db 73 KAKTLYDSTYHPDTEGEKMPFLGRMSAQPMNMTTGMLGLYRTLPGVFSHWFNOSFNA 132  
QY 133 IVNYSNRSQDPTITVRQLGTAVYSATGAVATAGLKSLLTKHLPLVGRFVPFAVAANA 192  
Db 133 VVNVTNRSGNSKATNERLFVSYCCATSGAMTVALGLNKVNVSHGLAARLVPPAIALAN 192  
QY 193 CINIPLMRQRELQVIGIPVADAGRLGYSVTAAGKQIFQVVISRICMAIPMAIPLIMD 252  
Db 193 AINIPMRSEASEGEMLEKDENDQLVGKSQKMAALSLAQVTLISRIAMAMPYVMTPTIWN 252  
QY 253 TLEKKDFLK 261  
Db 253 RITRTAYYR 261

RESULT 10  
SFX5\_HUMAN STANDARD; PRT; 340 AA.  
ID SFX5\_HUMAN  
AC Q8TD22;  
DT 15-SEP-2003 (Rel. 42, Created)  
DT 15-SEP-2003 (Rel. 42, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Sideroflexin 5.  
GN SFXN5.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.  
RX MEDLINE=22035372; PubMed=12039050;  
RA Lockhart P.J., Holtom B., Lincoln S., Hussey J., Zimprich A.,  
RA Gasser T., Waszolek Z.K., Hardy J., Farrer M.J.;  
RT "The human sideroflexin 5 (SFXN5) gene: sequence, expression analysis  
RT and exclusion as a candidate for PARK3.";  
RL Gene 285:229-237(2002).  
CC -!- FUNCTION: Potential iron transporter.  
CC -!- SUBCELLULAR LOCATION: Mitochondrial (By similarity).  
CC -!- TISSUE SPECIFICITY: Primarily expressed in the brain.  
CC -!- SIMILARITY: Belongs to the sideroflexin family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -



Db 271 IVMSMLEKTALQ 283

RESULT 12

SRE2\_CAEEL STANDARD; PRT; 600 AA.

ID SRE2\_CAEEL

AC 009273;

DT 01-NOV-1997 (Rel. 35, Created)

DT 15-SEP-2003 (Rel. 42, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Serpentine receptor class epsilon 2 (Sre-2 protein).

GN SRE-2 OR C41C4.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI\_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RA Burton J.;

RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.

RN [2]

RP REVISIONS.

RA Durbin R.;

RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).

CC -1- SIMILARITY: BELONGS TO THE C.ELEGANS RECEPTOR-LIKE PROTEIN SRE FAMILY.

-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

CC EMBL; Z48045; CAA88099.2; -.

DR PIR; T19873; T19873.

DR WormPep; C41C4.2; CE32323.

DR InterPro; IPR004686; Mtc.

DR InterPro; IPR004151; Sre.

DR Pfam; PF03820; Mtc; 1.

DR Pfam; PF03125; Sre; 1.

DR ProDom; PD006986; Mtc; 1.

DR TIGRFAMs; TIGR00798; mtc; 1.

KW Transmembrane; Multigene family.

FT TRANSMEM 129 149 POTENTIAL.

FT TRANSMEM 178 198 POTENTIAL.

FT TRANSMEM 269 289 POTENTIAL.

FT TRANSMEM 303 323 POTENTIAL.

FT TRANSMEM 380 400 POTENTIAL.

FT TRANSMEM 410 430 POTENTIAL.

FT TRANSMEM 466 486 POTENTIAL.

FT TRANSMEM 506 526 POTENTIAL.

SQ SEQUENCE 600 AA; 68517 MW; 41BAC8FB34C44CDS CRC64;

Query Match 32.2%; Score 432.5; DB 1; Length 600;

Best Local Similarity 40.0%; Pred. No. 1.2e-30;

Matches 100; Conservative 43; Mismatches 96; Indels 11; Gaps 6;

QY 16 EPRWDQSTFLGRARHFFVTDPNLLSGAQLASRNIVQNYRAGVTPGITEDQLWRAK 75

Db 44 EPRFPQDTFLGRYLHCLDIVDPRTLPAFNKKEESLELNSFKAGTAT-NVPDKSLWEAQ 102

QY 76 YVYDSAFHPDTGKLVLLGRMSAQVPMN-MTITGCMLETFYRKTPTVVFQWVNSFNATV 134

Db 103 KLKSAIHLHPDTGKVLPPFRMSGFVPPGWITYG-MLLPNPSWPTLLFQWQMNQSHNACV 161

QY 135 NYSNRSGDPTITVRQLGTAVVSATGAVATLGLKSLTK--HLPP-----LVGRFVPPAA 187

Db 162 NYANRNATQPPQLSKYIGAYGAATAACSGGLTYFIKASSLPPTRIIIQRFVPLPA 221

QY 188 VAAANCINIPLMRQRELQVGIPVAD-EAGORLGYSVTAQKQIFQVVISRICMAIPAMAI 246

Db 222 TSLASSLNVICMRWNELETGIQVYEXDGTGKVGVSVAQAQAVDTTWTVAFLVPLLM 281

QY 247 PPLIMDTLEK 256

Db 282 PRCIMPYLER 291

RESULT 13

SFX4\_MOUSE STANDARD; PRT; 313 AA.

ID SFX4\_MOUSE

AC Q925N1;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Sideroflexin 4.

GN SFXN4.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.

RX MEDLINE=21172735; PubMed=11274051;

RA Fleming M.D., Campagna D.R., Haslett J.N., Trenor C.C. III, Andrews N.C.;

RA "A mutation in a mitochondrial transmembrane protein is responsible for the pleiotropic hematological and skeletal phenotype of flexed-tail (f/f) mice."

RL Genes Dev. 15:652-657(2001).

CC -1- FUNCTION: Potential iron transporter.

CC -1- SUBCELLULAR LOCATION: Mitochondrial (By similarity).

CC -1- TISSUE SPECIFICITY: Largely restricted to kidney, brain and heart.

CC -1- SIMILARITY: Belongs to the sideroflexin family.

-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

CC EMBL; AF325263; AAK39431.1; -.

DR MGD; MGI:2137680; Sfxn4.

DR InterPro; IPR004686; Mtc.

DR Pfam; PF03820; Mtc; 1.

DR ProDom; PD006986; Mtc; 1.

KW Transmembrane; Iron transporter; Iron; Mitochondrion; Transmembrane.

FT TRANSMEM 87 107 POTENTIAL.

FT TRANSMEM 141 161 POTENTIAL.

FT TRANSMEM 175 191 POTENTIAL.

FT TRANSMEM 230 247 POTENTIAL.

FT TRANSMEM 269 289 POTENTIAL.

SQ SEQUENCE 313 AA; 35685 MW; ACB650B711A7552P CRC64;

Query Match 8.4%; Score 113; DB 1; Length 313;

Best Local Similarity 20.7%; Pred. No. 0.011;

Matches 56; Conservative 50; Mismatches 125; Indels 40; Gaps 8;

QY 10 LDINIOEPRWDQSTFLGRARHFFVTDPNLLSGAQLASRNIVQNYRAGVTPGITED 69

Db 1 MEPNLQFWISERQAFFRRFCQWMDLDEVNMFISIGSIEKSRQLFT-----TED 50

QY 70 QLWRKAYVD-----SAFHPDTGKLVLLGRMSAQVPMNMTITGCMLETFYRKT 117

Db 51 ---APKHYLDNQVIKDAWNKSLSTVHPDSSKLIPLFRPAAPLPTVAPMVFLIMPDTGI 107

QY 118 PTVVFQWVNSFNATVNSNRSGD---TPITVRQLGT-AVVSAT-TGAVATLGLKSLT 172

Db 108 KSIILLQSCLYGYTTAFNITNGNASYSHGFVERTLLGAGVFVSSFTIGLIPHLLQMKY-- 165



```

QY      173 KHLPP-----VGRFVPFAVAANNCINIPLMKRELOQVGIPVADAGORLGYVTAKQG 228
          || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      166 ----PLNNFWLKRTLPVIFLAQVSGMNVFASRSFENHRGIEVMDEKGHVGHGRRAKRKA 221
          || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

QY      229 IFQVISRICMAIPAMAI PPLIMDTLEKDF 259
          | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      222 IKDTAKSRAVLFGTSALAPELFIHFKRTRF 252
          | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

RESULT 14
GCSR_HUMAN
ID      GCSR_HUMAN        STANDARD;          PRT;       836 AA.
AC      Q99062;
DT      01-FEB-1995 (Rel. 31, Created)
DT      01-FEB-1995 (Rel. 31, Last sequence update)
DT      15-SEP-2003 (Rel. 42, Last annotation update)
DE      Granulocyte colony stimulating factor receptor precursor (G-CSF-R)
DE      (CD114 antigen).
GN      CSF3R OR GCSFR.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Placenta;
RX      MEDLINE=91079757; PubMed=2147944;
RA      Larsen A., Davis T., Curtis B.M., Gimpel S., Sims J.E., Cosman D.,
RA      Park L., Sorensen E., March C.J., Smith C.A.;
RT      "Expression cloning of a human granulocyte colony-stimulating factor
RT      receptor: a structural mosaic of hematopoietin receptor,
RT      immunoglobulin, and fibronectin domains.";
RL      J. Exp. Med. 172:1559-1570(1990).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Placenta;
RX      MEDLINE=91062348; PubMed=1701053;
RA      Fukunaga R., Seto Y., Mizushima S., Nagata S.;
RT      "Three different mRNAs encoding human granulocyte colony-stimulating
RT      factor receptor.";
RL      Proc. Natl. Acad. Sci. U.S.A. 87:8702-8706(1990).
RN      [3]
RP      SEQUENCE FROM N.A.
RC      MEDLINE=92091782; PubMed=1530796;
RA      Seto Y., Fukunaga R., Nagata S.;
RT      "Chromosomal gene organization of the human granulocyte colony-
RT      stimulating factor receptor.";
RL      J. Immunol. 148:259-266(1992).
RN      [4]
RP      SEQUENCE FROM N.A., AND VARIANTS THR-231; ASN-320; ARG-346; LYS-405;
RP      GLN-440; HIS-510; HIS-562 AND CYS-583.
RA      Rieder M.J., Armet T.Z., Carrington D.P., Ozuna M., Kuldanek S.A.,
RA      Rajkumar N., Toth E.J., Yi O., Nickerson D.A.;
RL      Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN      [5]
RP      DOMAINS STRUCTURE.
RX      MEDLINE=92007729; PubMed=1717255;
RA      Fukunaga R., Ishizaka-Ikeda E., Pan C.-X., Seto Y., Nagata S.;
RT      "Functional domains of the granulocyte colony-stimulating factor
RT      receptor.";
RL      EMBO J. 10:2855-2865(1991).
RN      [6]
RP      DISEASE.
RX      MEDLINE=94240159; PubMed=7514305;
RA      Dong F., Hoefeloot L.H., Schelen A.M., Broeders C.A., Meijer Y.,
RA      Veerman A.J., Touw I.P., Lowenberg B.;
RT      "Identification of a nonsense mutation in the granulocyte-colony-
RT      stimulating factor receptor in severe congenital neutropenia.";
RL      Proc. Natl. Acad. Sci. U.S.A. 91:4480-4484(1994).
RN      [7]
RP      STRUCTURE BY NMR OF 227-334.
RX      MEDLINE=97331327; PubMed=9187659;
RA      Yamasaki K., Naito S., Anaguchi H., Ohkubo T., Ota Y.;

```

```

RT "solution structure of an extracellular domain containing the WSxWS
RT motif of the granulocyte colony-stimulating factor receptor and its
RT interaction with ligand.";
RL Nat. Struct. Biol. 4:498-503 (1997).
RN [8]
RP 3D-STRUCTURE MODELING OF 125-331.
RX MEDLINE=98037802; PubMed=9368043;
RA Layton J.E., Iaria J., Smith D.K., Treutlein H.R.;
RT "identification of a ligand-binding site on the granulocyte colony-
RT stimulating factor receptor by molecular modeling and mutagenesis.";
RL J. Biol. Chem. 272:29735-29741 (1997).
CC -1- FUNCTION: RECEPTOR FOR GRANULOCYTE COLONY-STIMULATING FACTOR (G-
CC CSF). IN ADDITION IT MAY FUNCTION IN SOME ADHESION OR RECOGNITION
CC EVENTS AT THE CELL SURFACE.
CC -1- SUBUNIT: DIMER (PROBABLE).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. THE GCSFR-2 FORM,
CC WHICH LACKS THE TRANSMEMBRANE DOMAIN, MAY REPRESENT A SOLUBLE FORM
CC OF THE RECEPTOR.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Comment=Additional isoforms seem to exist. Experimental
CC confirmation may be lacking for some isoforms;
CC Name=1; Synonyms=GCSFR-1;
CC IsoId=Q99062-1; Sequence=Displayed;
CC Name=2; Synonyms=GCSFR-2;
CC IsoId=Q99062-2; Sequence=VSP_001674;
CC Name=3; Synonyms=GCSFR-3;
CC IsoId=Q99062-3; Sequence=VSP_001673;
CC Name=4; Synonyms=GCSFR-4, D7;
CC IsoId=Q99062-4; Sequence=VSP_001671, VSP_001672;
CC -1- TISSUE SPECIFICITY: ONE OR SEVERAL FORMS HAVE BEEN FOUND IN
CC MYELOGENOUS LEUKEMIA CELL LINE KG-1, LEUKEMIA U937 CELL LINE, IN
CC BONE MARROW CELLS, PLACENTA, AND PERIPHERAL BLOOD GRANULOCYTES.
CC THE GCSFR-2 FORM HAS BEEN FOUND ONLY IN LEUKEMIA U937 CELLS. THE
CC GCSFR-3 FORM IS HIGHLY EXPRESSED IN PLACENTA.
CC -1- DISEASE: Defects in CSF3R might be a cause of severe congenital
CC neutropenia (SCN) in some patients.
CC -1- SIMILARITY: Contains 5 fibronectin type III domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like C2-type domain.
CC -1- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.
CC -1- DATABASE: NAME=PROW, NOTE=CD guide CD114 entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd114.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.itsb-stb.ch/announce/
CC or send an email to license@itsb-stb.ch).
CC -----
DR EMBL; X55721; CAA39253.1; -.
DR EMBL; X55720; CAA39252.1; -.
DR EMBL; S71484; AAB20660.1; -.
DR EMBL; M59818; AAA63176.1; -.
DR EMBL; M59819; AAA63177.1; -.
DR EMBL; M59820; AAA63178.1; -.
DR EMBL; AY148100; AAN05790.1; -.
DR PIR; B38252; B38252.
DR PIR; C38252; C38252.
DR PIR; JH0329; JH0329.
DR PDB; 1AZ7; 28-JAN-98.
DR Genew; HGNC:2439; CSF3R.
DR MIM; 138971; -.
DR GO; GO:0005887; C:Integral to plasma membrane; TAS.
DR GO; GO:0004872; F:receptor activity; TAS.
DR GO; GO:0006952; P:defense response; TAS.
DR InterPro; IPR002996; CRIA.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003529; Hemtopoptn_L_F2.
DR Pfam; PF00041; fn3; 3.
DR SMART; SM00060; FN3; 2.
DR PROSITE; PS01353; HEMATOPO REC L_F2; 1.

```





FT DISULFID 243 295 POTENTIAL.  
FT CARBOHYD 44 44 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 70 70 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 152 152 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 216 216 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 284 284 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 292 292 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 321 321 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT LIPID 321 321 GPI-ANCHOR (POTENTIAL).  
SQ SEQUENCE 344 AA; 37998 MW; CBB39BE53B33B224 CRC64;

Query Match. 6.7%; Score 89.5; DB 1; Length 344;  
Best Local Similarity 24.8%; Pred. No. 1.4;  
Matches 52; Conservative 28; Mismatches 95; Indels 35; Gaps 10;

QY 2 ESKMGEPLDINIQEPRW-DOSTFLGRARHFTVTDPNLLSGAQLASRNI--VQNYR 58  
Db 51 ESATLRCTIDNRVTRVAMLNRSSTILYAGNDKWCL-DPRVLLSNTQTQYSIEIQNVVYD 109  
QY 59 AGVVTGITEDQLMRAKYVYDSAFHPDTGKXVVLIGRMSAQV-----PMNMTIT 107  
Db 110 EGPYTCVQTDN-----HPKT-SRVHLIVQVSPKIVEISSDISINEGNNISLT 156  
QY 108 GCMLEFYRKPTVFVQWVNSFNALVNSNRSGDPTIVRQLGTAYVSATGAVATAIG 167  
Db 157 -CIAT-GRPEPTVT-WRHISPKAVGFVSEDEYLEIQGITREQSGEYECASASNDVAAPVR 213  
QY 168 LKSLTKHLPLV---GRFVPPAAVAAANC 193  
Db 214 RVNVTVNYPYISEAKGTGVPVGQKGTLC 243

Search completed: February 3, 2004, 09:13:15  
Job time : 19 secs



```

QY 121 VFQWVNSFNAIVNYSNRSGDPTITVROLGTAVSATGAVATAGLKSITKHLPLV 180
Db 121 VFQWVNSFNAIVNYSNRSGDPTITVROLGTAVSATGAVATAGLKSITKHLPLV 180
QY 181 RFVFPAAVAANCINIPLMRQRELQVGPVADAGQRLGYSVTAKOGIFQVVISRICMA 240
Db 181 RFVFPAAVAANCINIPLMRQRELQVGPVADAGQRLGYSVTAKOGIFQVVISRICMA 240
QY 241 IPAMAIPLIMDTLEKDFLK 261
Db 241 IPAMAIPLIMDTLEKDFLK 261

RESULT 2
Q8C1Z2 PRELIMINARY; PRT; 288 AA.
ID Q8C1Z2
AC Q8C1Z2;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Sideroflexin 3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK089985; BAC41029.1; -.
SQ SEQUENCE 288 AA; 31630 MW; E388501B49D6EC2A CRC64;

Query Match 79.1%; Score 1063.5; DB 11; Length 288;
Best Local Similarity 82.9%; Pred. No. 5.1e-89;
Matches 213; Conservative 6; Mismatches 5; Indels 33; Gaps 1;

QY 5 MGELPLDINIOEPRWDOSTFLGRARHFTVTDPRNLLSGAOLEASRNIVQNYRAGVATP 64
Db 1 MGDLPLNINIOEPRWDOSTFLGRARHFTVTDPRNLLSGAOLEASRNIVQNYRAGVATP 60
QY 65 GITEDQLMRAKYVYDSAFHPDTGKVLIGRMSAQVPMNMTITGCMILTFYRKTPTVFWQ 124
Db 61 GLTEDQLMRAKYVYDSAFHPDTGKVLIGRMSAQVPMNMTITGCMILTFY----- 110
QY 125 WVNQSFNAIVNYSNRSGDPTITVROLGTAVSATGAVATAGLKSITKHLPLVGRFVP 184
Db 111 -----ROLGTAVSATGAVATAGLKSITKHLPLVGRFVP 147
QY 185 FAAVAAANCINIPLMRQRELQVGPVADAGQRLGYSVTAKOGIFQVVISRICMAIPAM 244
Db 148 FAAVAAANCINIPLMRQRELQVGPVADAGQRLGYSVTAKOGIFQVVISRICMAIPAM 207
QY 245 AIPLIMDTLEKDFLK 261
Db 208 AIAPVIMNTLEKDFLK 224

RESULT 3
Q9VN13 PRELIMINARY; PRT; 321 AA.
ID Q9VN13;
AC Q9VN13;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE CG11739 protein (AT24389p) (RH48017p).
GN CG11739.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

```

```

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieres S., Fleischmann W.,
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclele J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svrtkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise B.,
RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunco J., Paclele J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise B.,
RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunco J., Paclele J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [9]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [10]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [11]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [12]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [13]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [14]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [15]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [16]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [17]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [18]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [19]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [20]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [21]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [22]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [23]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [24]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [25]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [26]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [27]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [28]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [29]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [30]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [31]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [32]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [33]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [34]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [35]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [36]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [37]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [38]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [39]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [40]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [41]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [42]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [43]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [44]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [45]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [46]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [47]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [48]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [49]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [50]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [51]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [52]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [53]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [54]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [55]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [56]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [57]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [58]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [59]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [60]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [61]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [62]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [63]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [64]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [65]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [66]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [67]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [68]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [69]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [70]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [71]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [72]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [73]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [74]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [75]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [76]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [77]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [78]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [79]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [80]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [81]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [82]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [83]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [84]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [85]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [86]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [87]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [88]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [89]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [90]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [91]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [92]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [93]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [94]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [95]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [96]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [97]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [98]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [99]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [100]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.

```





OS *Drosophila melanogaster* (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
[1]  
RN  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkeley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Talati M., Kalush F., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Splier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svrtkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,  
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,  
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
RA Stapleton M., Strong R., Svrtkas R., Tector C., Tyler D.,  
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
RT "Sequencing of *Drosophila melanogaster* genome.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,  
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,  
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;

```

RT "Annotation of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FLYBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003518; AAF49194.2; -
DR FLYBase; FBgn0036843; CG6812.
DR InterPro; IPR004686; Mtc.
DR Pfam; PF03820; Mtc; 1.
DR TIGRFAMs; TIGR00798; mtc; 1.
SQ SEQUENCE 327 AA; 36495 MW; D3D235768ED50200 CRC64;

Query Match
Best Local Similarity 45.8%; Score 615.5; DB 5; Length 327;
Matches 128; Conservative 46; Mismatches 78; Indels 1; Gaps 1

QY 5 MGELPLDINIOEPRWDSTFLGRARHFTVTDPRNLLSGAQLASRNIVQNYRAGVTP 64
DB 1 MSQVSTLIDVDKPLFDLSTFAGRFPQYPAWMTDPRTVVSSDRLLBAKAVERYRKGDQSP 60
QY 65 GITEDQLMRKAVYDSAPHDPTGEKVVLIGRMSAQVPMAMTTTCMLTFYRKTPTVVFWQ 124
DB 61 PLKPEEVHYNMKNLNSAFHPDTGELQNPCGRMSFQVPGGMLITGCMIAFYRTVPAVLMQ 120
QY 125 WVNQSFNAIVNYSNRSGDPTITVRQLGTAYVSATTGAVATAGLKL-TKHLPLVGRFV 183
DB 121 FINQSFNAVNTYRNANSPTSVTQLGVAYSATTSALVAALGCKNYSKATPLPQRFV 180
QY 184 PFAAFAAANCINIPLMQRELQVGI PVADAGQRLGYSVTAAGKQIFQVVISRICMAIPA 243
DB 181 PFAAFAAANFVNIPLMRQNEIINGIEVKNDGDVVVGQSRLLAIKIGIEVVSRIMAAAPG 240
QY 244 MAIPPLIMDTLEK 256
DB 241 MLVLPLIMERLEK 253

RESULT 7
ID Q8S296 PRELIMINARY; PRT; 327 AA.
AC Q8S296;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE R31111p.
GN CG6812.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RA Stapleton M., Brokstein P., Hong L., Agbayan A., Carlson J.,
RA Champagne M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunco J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY071029; AAL48651.1; -
DR FLYBase; FBgn0036843; CG6812.
DR InterPro; IPR003583; HHH_1.
DR InterPro; IPR004686; Mtc.
DR Pfam; PF03820; Mtc; 1.
DR SMART; SM00278; HhH1; 1.
DR TIGRFAMs; TIGR00798; mtc; 1.

```

SQ	SEQUENCE	327 AA;	36468 MW;	AED2357693D50B92 CRC64;
	Query Match	45.7%;	Score 614.5;	DB 5; Length 327;
	Best Local Similarity	50.6%;	Pred. No. 7e-48;	
	Matches 128;	Conservative 45;	Mismatches 79;	Indels 1; Gaps 1
QY	5	MGELPLDINIQEPRWDQSTFLGRARHFTVTDPRNLLLSGAQLSEASRNIVQNRYAGVTP	64	
	:::	:::	:::	:::
Db	1	MSQVSTLIDVDKPLFDLSTFAGRFOYFAWMTDPRTVVVSSDRLLLEAKMVERYYRKGDQP	60	
QY	65	GITEDQLMRKATVYDSAFHPDPTGEKVLLIGMSAQVPMMNTITGCMLTFYRKTPTVFVWQ	124	
	::::	::::	::::	::::
Db	61	PLKPEEVHYNMKLINSAFHPDYGELONFCGRMSFQVPGCMLITGGMLAFYRTVPAVLWQ	120	
QY	125	WNQSFNAIVNYSNRSGDPTITVAQLGTAIVSATGAVATAALGLKS-LTKHLPLVGRFV	183	
	:::	:::	:::	:::
Db	121	FINQSFNAVNVNYTNRRANSPTSVTQGVAVYSATTSAALVAAGCKMYWSKKATPPLFORFV	180	
QY	184	PFAAVALAACINIPLMQRQLQVGIPVADEAGQRLGYSVTAAKGQIFQVVISRICMAIPA	243	
Db	181	PFAAVALAANFVNIPLMQRQNEIINGIEVKSDDGVVVGQSRLLAIKGIETVVVSRIAMAAPG	240	
QY	244	MAIPPLIMDTLEK	256	
	:			
Db	241	MLVLPIMERLEK	253	

ID	Accession	Source	Length	Score	DB	Pred.	Matches	Conservative	Mismatches	Indels	Gaps
017863	PRELIMINARY;	PRT;	349 AA.								
AC	017863;										
DT	01-JAN-1998 (TREMBLrel. 05, Created)										
DT	01-OCT-2001 (TREMBLrel. 18, Last sequence update)										
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)										
DE	F37H8.4 protein.										
GN	F37H8.4.										
OS	Caenorhabditis elegans.										
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;										
OX	Rhabditidae; Peloderinae; Caenorhabditis.										
NCBI	TaxID=6239;										
RP	SEQUENCE FROM N.A.										
RA	Gregory J.;										
RL	Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.										
RP	[2]										
RP	SEQUENCE FROM N.A.										
RX	MEDLINE=99069613; PubMed=9851916;										
RA	none;										
RT	"Genome sequence of the nematode C.elegans: A platform for										
RT	investigating biology.";										
RL	Science 282:2012-2018(1998).										
DR	EMBL: Z81534; CAB04347.2; -.										
DR	WormPep; F37H8.4; CE28302.										
DR	InterPro; IPR004686; Mtc.										
DR	Pfam; PF03820; Mtc; 1.										
DR	TIGRFAMS; TIGR00798; mtc; 1.										
SD	SEQUENCE 349 AA; 39351 MW; 72CA25828404EF82 CRC64;										
Query Match											
Best Local Similarity				43.3%;	Score 582.5;	DB 5;	Length 349;				
Matches 121; Conservative				43.4%;	Pred. No. 6.5e-45;						
				47;	Mismatches 86;	Indels 25;	Gaps 2				
QY	1 MESKMGELPLDINIQEPKWDQSTFLG-----	PARHFTVTDP	37								
DB	1 MSELVSNLKTTPDISKPQMSQRTYYGKNSGCGQKVPAKRLFRWDVSESRVRHFTLTNP		60								
QY	38 RNLLLSGAQLASRNIVQNYRACGVVTPGITEDQLMRAKYVYDSAFHPDTGEKVVILGRMS		97								
DB	61 LTLTSSVARQEQCRQIVLDYKNGKVSPTLTIVSELMKAKTLIDSTYHPDTEKMFLLGRMS		120								
QY	98 AQVPMNTITGCMITFYRKTPVYFWQWVNSFNAIVNYSNRSGDTPITVRQLGTAYVSA		157								
DB	121 AQPMANMLINGMILLSYRTFPGVVFESHWINQSFNNAVNTNRSNGSKTSNERLILSYSCA		180								

[illegible]

```

RESULT 9
ID Q09972 PRELIMINARY; PRT; 319 AA.
AC Q09972;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical 35.3 kDa protein.
GN C14F5.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditiidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; Pubmed=9851916;
RA Waterston R.;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Mix P.;
RT "The sequence of C. elegans cosmid C14F5.";
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
EMBL: U29082; AAA68404.3; -.
DR WormPep; C14F5.4; CE30370.
DR InterPro; IPR004686; MTC.
DR Pfam; PF03820; MTC; 1.
DR TIGRFAMs; TIGR00798; mtc; 1.
KW Hypothetical protein.
SQ SEQUENCE 319 AA; 35323 MW; B400FEA731B27F93 CRC64;

```

Query Match	37.8%	Score 508.5	DB 5	Length 319
Best Local Similarity	44.4%	Pred. No. 3.3e-38		
Matches 108	Conservative 39	Mismatches 93	Indels 3	Gaps 2
Qy	16	EPKRDQSTFLGRARHFFVTDPBRNLLSGAQLKASRNIVQNYRAGVTTGITEDQLWRAK	75	
Dp	7	QPRYDQSTFYGRLRHFAGMTDPLIAFSSTTELLITASELMQCKREKKVPA-TLEELHRSQ	65	
Qy	76	YVYDSAFHPDTEGEKVLLIGMSAQVPMNMTITGCMLTFYRKTPTFVFWQWVNSFNAIVN	135	
Dp	66	RLQCSAFHPDTEGELQNFAGRMCFNVWGTMLCGAMMIWYKSTPAVIFWQWANSFNALVN	125	
Qy	136	YSNRSGDPTITVRQLGTAYVSATGAVATALGLKS--LTKHLPLVGRFVFPAAVAANC	193	
Dp	126	YTRNNAKSTLTTKDLVVSYSTAVSGALAMAIGLKTYFAKKQSSPLAQRLLVPLGAVAVANA	185	
Qy	194	INIPLMRQRELQVGIVPAVDEAGORLGYSVTAAKQGIPOVVISRICMAIPMAIPLPIMDT	253	
Dp	186	INIPMRQNELKEGMYTVDADGNNVGVSRLLAAKAISLVLSRNIIIVAPCMILTPVIMEG	245	
Qy	254	LEK	256	
Dp	246	LNK	248	





Db 211 ASANICNVLMRYGELBEGIDVLDADGNLVGSSKIAARHALLETALTTRVLPMPILVLP 270  
Qy 249 LIMDTLEKKDFLK 261  
Db 271 IVMSMLEKTALLQ 283

RESULT 13

Q8BGG3 PRELIMINARY; PRT; 342 AA.  
ID Q8BGG3  
AC Q8BGG3;  
DT 01-MAR-2003 (Tremblrel. 23, Created)  
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
DE Sideroflexin 5.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RE SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs."  
RL Nature 420:563-573(2002).  
DR EMBL; AK036285; BAC29372.1; -.  
DR EMBL; AK082073; BAC38403.1; -.  
SQ SEQUENCE 342 AA; 37328 MW; 9BDB8261D5EF79D9 CRC64;

Query Match 36.0%; Score 484.5; DB 11; Length 342;  
Best Local Similarity 38.3%; Pred. No. 5.8e-36;  
Matches 97; Conservative 50; Mismatches 99; Indels 7; Gaps 2;

Qy 16 EPRNDSTFLGRARHFTVTDPNLLSGAQLBASRNIVQNYRAGVTPGITEDQLWRAX 75  
Db 31 KPRFOQTSFYGRFRHFLDIDPRITLVTEKRLREAVQLLEDYKHGTLRPGVTNEQLWSAQ 90  
Qy 76 YVYDSAFHPDTGEKVLIGRMSAQVPMNMTITGCMLEFYRKPTVVFQWVNOSFNAYN 135  
Db 91 KIKQAILHPDTNEKIFMPFRMSGYIPFGTPIVGLLPNQTLASTVFWQMLNOSHNA CVN 150  
Qy 136 YSNRSGDPTITVRQLGTAYVSATTGAVALGLKSL---TKHLPP--LVGRFVPFAAV 188  
Db 151 YANRNATKPSPAKFIQGYLGAVISAVSIAGLVNLVQKANKFTPATRLLVQRFVPPPAV 210  
Qy 189 AAANCINIPLMRQRELQVGPVADAGQRLGYSVTAAGQIFQVVISRICMAIPAMAIPP 248  
Db 211 ASANICNVLMRYGELBEGIDVLDADGNLVGSSKIAARHALLETALTTRVLPMPILVLP 270  
Qy 249 LIMDTLEKKDFLK 261  
Db 271 IVMSMLEKTALLQ 283

RESULT 14

Q8BRQ9 PRELIMINARY; PRT; 302 AA.  
ID Q8BRQ9  
AC Q8BRQ9;  
DT 01-MAR-2003 (Tremblrel. 23, Created)  
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
DE Sideroflexin 5.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RE SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Cortex;

RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs."  
RL Nature 420:563-573(2002).  
DR EMBL; AK043706; BAC31624.1; -.  
SQ SEQUENCE 302 AA; 32783 MW; 650F6A997043EF26 CRC64;

Query Match 35.9%; Score 483.5; DB 11; Length 302;  
Best Local Similarity 38.6%; Pred. No. 6e-36;  
Matches 96; Conservative 50; Mismatches 96; Indels 7; Gaps 2;

Qy 16 EPRNDSTFLGRARHFTVTDPNLLSGAQLBASRNIVQNYRAGVTPGITEDQLWRAX 75  
Db 31 KPRFOQTSFYGRFRHFLDIDPRITLVTEKRLREAVQLLEDYKHGTLRPGVTNEQLWSAQ 90  
Qy 76 YVYDSAFHPDTGEKVLIGRMSAQVPMNMTITGCMLEFYRKPTVVFQWVNOSFNAYN 135  
Db 91 KIKQAILHPDTNEKIFMPFRMSGYIPFGTPIVGLLPNQTLASTVFWQMLNOSHNA CVN 150  
Qy 136 YSNRSGDPTITVRQLGTAYVSATTGAVALGLKSL---TKHLPP--LVGRFVPFAAV 188  
Db 151 YANRNATKPSPAKFIQGYLGAVISAVSIAGLVNLVQKANKFTPATRLLVQRFVPPPAV 210  
Qy 189 AAANCINIPLMRQRELQVGPVADAGQRLGYSVTAAGQIFQVVISRICMAIPAMAIPP 248  
Db 211 ASANICNVLMRYGELBEGIDVLDADGNLVGSSKIAARHALLETALTTRVLPMPILVLP 270  
Qy 249 LIMDTLEKK 257  
Db 271 IVMSMLEKQ 279

RESULT 15

Q12029 PRELIMINARY; PRT; 327 AA.  
ID Q12029  
AC Q12029;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)  
DE Chromosome XV reading frame ORF YOR271C.  
GN YOR271C.  
OS Saccharomyces cerevisiae (Baker's Yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;  
RN [1]  
RE SEQUENCE FROM N.A.  
RA Jauniaux J.C., Poirey R.;  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RE SEQUENCE FROM N.A.  
RA Cheret G., Sor F.;  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RE SEQUENCE FROM N.A.  
RA MIPS;  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RE SEQUENCE FROM N.A.  
RC STRAIN=S288C;  
RA Cheret G., Bernardi A., Sor F.J.;  
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Z75179; CAA99495.1; -.  
DR EMBL; X89633; CAA61777.1; -.  
DR SGD; S0005797; YOR271C.  
DR InterPro; IPR004686; Mtc.  
DR Pfam; PF03820; Mtc; 1.  
DR TIGRFAMs; TIGR00798; mtc; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 327 AA; 35414 MW; FD175626E63B8619 CRC64;

Query Match		34.5%;	Score 464.5;	DB 3;	Length 327;
Best Local Similarity		40.6%;	Pred. No. 3.7e-34;		
Matches 104;		Conservative 41;	Mismatches 102;	Indels 9;	Gaps 2;
QY	8	LPIDINIQEPRWDSTFLGRARHFTVTDPRLILLSGAQLASRNIVQNYRAGVTPGIT	67		
Db	5	VPGPIDLPESRYDLSTYWGRI RHCAEISDPTMLLTTEKOLAHAREIISAYRHGEIKE--T	62		
QY	68	EDQLWRAKYVYDSAFHPDTGKVVLLIGMSAQVPMNMITITGCMLTFRKTPPTVFWQWVN	127		
Db	63	TPEFWRAKKQLDSTVHPDTGKTVLLPFRMSSNVL SNLVTVGM LTPGLGTAGTIVFWQMAN	122		
QY	128	QSFNALVNSNRSGDPTITVRQLGTAYVSATTGAVATAG-----LKSILTKHLPLPLVG	180		
Db	123	QSLNVAVNSANANKSHPMSTSQLLTNYAAAVTASCGVALGNNLVPR LKNISPHSKLILG	182		
QY	181	RFVPEFAVAANAANCINIPLMRQRELQVGIPVADEAGRLGYSVTAAKQGI FQVVISRICMA	240		
Db	183	RLVPEFAAVVSAGIVNVFLMRGNEIRKGISVFDSDNGDEVGSKSKAAFMVAGETALSRVTNA	242		
QY	241	IPAMAI PPLIMDTLEK	256		
Db	243	TPTMVI PPLILVRLQR	258		

Search completed: February 3, 2004, 09:14:43  
Job time : 44 secs

QY 241 IPMAIPLIMDTLEKDFLK 261  
Db 241 IPMAIPLIMDTLEKDFLK 261

RESULT 2

US-09-867-550-1344  
; Sequence 1344, Application US/09867550  
; Patent No. US20020082206A1  
; GENERAL INFORMATION:  
; APPLICANT: Leach, Martin D.  
; APPLICANT: Mehraban, Foad,  
; APPLICANT: Conley, Pamela  
; APPLICANT: Law, Debbie  
; APPLICANT: Topper, James  
; TITLE OF INVENTION: No. US20020082206A1 Polynucleotides from Atherogenic Cells and  
; FILE REFERENCE: 21402-013 (Cura-313)  
; CURRENT APPLICATION NUMBER: US/09/867,550  
; PRIOR FILING DATE: 2001-09-20  
; NUMBER OF SEQ ID NOS: 2125  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1344  
; LENGTH: 266  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-867-550-1344

Query Match  
Best Local Similarity 100.0%; Score 1345; DB 9; Length 266;  
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MESKMGELPLDINIQEPRWDOSTFLGRARHFTVTDPNNLLSGAOLEASRNIVONYRAG 60  
Db 1 MESKMGELPLDINIQEPRWDOSTFLGRARHFTVTDPNNLLSGAOLEASRNIVONYRAG 60  
QY 61 VTPGTEEDOLMRKAYVYDSAFHPDTGKVVILGRMSAQPMMNTITGCMLEFYRKTPTV 120  
Db 61 VTPGTEEDOLMRKAYVYDSAFHPDTGKVVILGRMSAQPMMNTITGCMLEFYRKTPTV 120  
QY 121 VFMQWVNSFNAINVNSRSGDPTITVRQLGTAVYSATTGAVATAGLKLTKLPLVPG 180  
Db 121 VFMQWVNSFNAINVNSRSGDPTITVRQLGTAVYSATTGAVATAGLKLTKLPLVPG 180  
QY 181 RFVPEFAVAANCINIPLMRQRELQVGPVADAGORLGYSVTAAKQGFQVVISRICMA 240  
Db 181 RFVPEFAVAANCINIPLMRQRELQVGPVADAGORLGYSVTAAKQGFQVVISRICMA 240  
QY 241 IPMAIPLIMDTLEKDFLK 261  
Db 241 IPMAIPLIMDTLEKDFLK 261

RESULT 3

US-09-990-415A-8  
; Sequence 8, Application US/09990415A  
; Patent No. US20020165182A1  
; GENERAL INFORMATION:  
; APPLICANT: Pharmacia AB  
; TITLE OF INVENTION: Protein Cluster I  
; FILE REFERENCE: 00349  
; CURRENT APPLICATION NUMBER: US/09/990,415A  
; NUMBER OF SEQ ID NOS: 2001-11-21  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 8  
; LENGTH: 322  
; TYPE: PRT  
; ORGANISM: human  
; FEATURE:  
NAME/KEY: misc\_feature

LOCATION: (25)..(25)  
; OTHER INFORMATION: Xaa=A,T,G or C  
US-09-990-415A-8

Query Match  
Best Local Similarity 81.0%; Score 1089; DB 10; Length 322;  
Matches 201; Conservative 28; Mismatches 27; Indels 0; Gaps 0;

QY 6 GELPLDINIQEPRWDOSTFLGRARHFTVTDPNNLLSGAOLEASRNIVONYRAGVTPG 65  
Db 3 GELPPNINIKPRWDOSTFLGRARHFTVTDPNNLLSGAOLEASRNIVONYRAGVTPG 65  
QY 66 ITEDOLMRKAYVYDSAFHPDTGKVVILGRMSAQPMMNTITGCMLEFYRKTPTVFWQ 125  
Db 63 LLENELMRKAYVYDSAFHPDTGKVVILGRMSAQPMMNTITGCMLEFYRKTPTVFWQ 125  
QY 126 VNSFNAINVNSRSGDPTITVRQLGTAVYSATTGAVATAGLKLTKLPLVGRFVP 185  
Db 123 INQSFNAVYVYNSRSGDPTITVRQLGTAVYSATTGAVATAGLKLTKLPLVGRFVP 185  
QY 186 AAVAAANCINIPLMRQRELQVGPVADAGORLGYSVTAAKQGFQVVISRICMAIPMA 245  
Db 183 AAVAAANCINIPLMRQRELQVGPVADAGORLGYSVTAAKQGFQVVISRICMAIPMA 245  
QY 246 IPPIMNTLEKDFLK 261  
Db 243 IPPIMNTLEKDFLK 258

RESULT 4

US-10-014-338-2  
; Sequence 2, Application US/10014338  
; Publication No. US20030092614A1  
; GENERAL INFORMATION:  
; APPLICANT: Herath, et al.  
; TITLE OF INVENTION: ADPI-41, A NOVEL PROTEIN ISOLATED FROM BRAIN TISSUE HOMOGENATE A  
; FILE REFERENCE: 9195-077  
; CURRENT APPLICATION NUMBER: US/10/014,338  
; PRIOR FILING DATE: 2002-05-01  
; NUMBER OF SEQ ID NOS: 10/014,338  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 322  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-014-338-2

Query Match  
Best Local Similarity 81.0%; Score 1089; DB 15; Length 322;  
Matches 201; Conservative 28; Mismatches 27; Indels 0; Gaps 0;

QY 6 GELPLDINIQEPRWDOSTFLGRARHFTVTDPNNLLSGAOLEASRNIVONYRAGVTPG 65  
Db 3 GELPPNINIKPRWDOSTFLGRARHFTVTDPNNLLSGAOLEASRNIVONYRAGVTPG 65  
QY 66 ITEDOLMRKAYVYDSAFHPDTGKVVILGRMSAQPMMNTITGCMLEFYRKTPTVFWQ 125  
Db 63 LLENELMRKAYVYDSAFHPDTGKVVILGRMSAQPMMNTITGCMLEFYRKTPTVFWQ 125  
QY 126 VNSFNAINVNSRSGDPTITVRQLGTAVYSATTGAVATAGLKLTKLPLVGRFVP 185  
Db 123 INQSFNAVYVYNSRSGDPTITVRQLGTAVYSATTGAVATAGLKLTKLPLVGRFVP 185  
QY 186 AAVAAANCINIPLMRQRELQVGPVADAGORLGYSVTAAKQGFQVVISRICMAIPMA 245  
Db 183 AAVAAANCINIPLMRQRELQVGPVADAGORLGYSVTAAKQGFQVVISRICMAIPMA 245  
QY 246 IPPIMNTLEKDFLK 261  
Db 243 IPPIMNTLEKDFLK 258